

PERCEPTUAL AGREEMENT BETWEEN MULTI-RATER FEEDBACK SOURCES

IN THE FEDERAL BUREAU OF INVESTIGATION

David Sandt Corderman

Dissertation submitted to the faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Human Development

Albert Wiswell, PhD, Chair

Marcie Boucouvalas, PhD

Letitia Combs, EdD

John Jarvis, PhD

Eric McCollum, PhD

April 16, 2004

Falls Church, Virginia

Key Words: Multi-rater Feedback, Leadership Development

PERCEPTUAL AGREEMENT BETWEEN MULTI-RATER FEEDBACK SOURCES  
IN THE FEDERAL BUREAU OF INVESTIGATION

David Sandt Corderman

(ABSTRACT)

The use of multi-rater feedback as a way to analyze perceptions within the context of job performance and leadership in the Federal Bureau of Investigation (FBI) was examined. Research in this domain is notable as this type of evaluation is now being done with regularity in the private sector and is starting to be utilized more extensively in the public sector, but is still being used to a limited extent in law enforcement. The path of this research examined differences between self-assessments and assessments of others (peers and subordinates) in dimensions of leadership as measured by the same multi-rater instrument at two points in time. This research effort made use of a multi-rater survey instrument called the "Leadership Commitments and Credibility Inventory System (LCCIS)," designed by Keilty, Goldsmith, and Company, which is used in multiple industries and was expanded to capture characteristics considered important to FBI leaders.

Results showed high ratings on a five point Likert scale as indicated by mean averages of self and others. Additionally, Z scores, t tests and ANCOVA indicated that FBI supervisors did not overestimate their leadership, as indicated by (1) an overall leadership measure at time two compared to time one, (2) a greater perceptual agreement between others and self existing on second multi-rater assessments than on the initial assessments, and (3) any statistical differences of means in all measured categories at time two versus time one. Various subcategories of the assessment showed a mixture of non-statistically significant results and that subordinates and peers perceived leaders differently. Further, analysis of two unique dimensions of the LCCIS, "Manage Diversity" and "Build Public Trust" showed exceptionally high results.

The implications of the present research are that leadership in the FBI, as measured by different dimensions, is strong. Yet, there is no evidence that leaders or others in this organization change their perceptions over time. These findings may point to the need for multi-rater instruments to be used in concert with personal development plans in order to improve the perception of leadership.

## Acknowledgements

I wish to thank a number of people for their support during my formal educational process. First, and most importantly, is my lovely wife, Louise. She has stood by me for more than 23 years, sustaining me in a multitude of ways including my academic pursuits which have consumed my attention for half of our marriage. Next, I want to thank my two children for their love and support.

I would also like to thank Dr. John Campbell, Dr. Lois Knowles, Dr. Marcie Boucouvalas, and Michael Ferrence, Jr. who originally had the foresight to see the value of the Adult Learning and Human Resource Development Program at Virginia Tech, and initiated the program of which I was a part. Jeffrey Higginbotham also deserves credit for supporting this effort. This program, and the three (and counting) that followed it, produced a significantly more sophisticated teaching corps at the FBI and DEA Academies. Thank you each for all you did to make these programs possible.

Further, I want to thank Dr. John Greene for all his assistance and guidance, and the members of my cohort and working groups who have sustained me during this process. All of you have nurtured and supported me throughout this effort, and I could not have done it without you. Additionally, special recognition must also go to Dr. Milton Goldsamt and Dr. Larry “Nick” Nicholson for pulling me to the finish line.

Finally, I wish to express my gratitude to the members of my committee. You allowed me to pursue an intellectual challenge, one that I enjoyed immensely. Special thanks must be given to committee members’ Dr. John Jarvis and Dr. Albert Wiswell, both of whom met with me on several occasions, steering me through the tumultuous process of the dissertation challenge.

## Table of Contents

Chapter One .....	1
Introduction .....	1
Overview .....	1
Conflicting Opinions on the Use of Multi-rater Feedback .....	2
360-Degree Feedback and Development .....	3
Multi-rater Instrument Used for this Study .....	4
The Federal Bureau of Investigation .....	4
Adult Learning and Human Resource Development .....	6
Leadership .....	6
The Issues .....	7
Research Questions .....	8
Definition of Terms .....	9
Outline of the Remainder of the Study .....	10
Chapter Two .....	11
Review of the Literature .....	11
Overview .....	11
Foundation .....	11
Perception .....	12
<i>Self-Perception Versus Person Perception</i> .....	13
<i>Other Biases and Problems in the Perception Process</i> .....	14
<i>Perception Change Theories</i> .....	15
Multi-rater Feedback .....	16
<i>Process and Purpose</i> .....	17
<i>Differences from Traditional Feedback Process</i> .....	18
<i>Reasons Organizations Use Multi-rater Feedback</i> .....	18
<i>Varying Uses of Multi-rater Feedback</i> .....	19
<i>Multi-rater Feedback from Direct Reports, Peers and Self-Reports.</i> ...	21
<i>Goal Setting and Multi-rater Feedback</i> .....	21
<i>Multi-rater Feedback and Law Enforcement</i> .....	22
Leadership .....	23
<i>Importance of Leadership</i> .....	24
<i>Leadership as a Performing Art</i> .....	24
<i>Developing Leaders</i> .....	24
<i>Leaders as Learners</i> .....	25
<i>Leadership and Reflection</i> .....	26
<i>Leadership and Feedback</i> .....	27
Perceptual Agreement Between Multi-rater Feedback Sources .....	27
<i>Agreement Between Self-assessment and Others' Assessments</i> .....	28
<i>Effect of Feedback on Self-Rating Perception Versus Others'</i> <i>Assessments</i> .....	29
Summary .....	29

Chapter Three .....	31
Methodology and Procedures .....	31
Introduction .....	31
Overview .....	31
Context... .....	31
Setting ... .....	33
Participants .....	37
Raters .....	37
Instrument.....	38
Procedures .....	41
Research Design and Analysis .....	42
Conclusion .....	44
Chapter Four .....	46
Research Findings .....	46
Introduction .....	46
Purpose... .....	46
Instrument Summary .....	47
Data Analysis.....	47
<i>Hypotheses</i> .....	49
<i>Hypothesis 1</i> .....	49
<i>Hypothesis 2</i> .....	52
<i>Hypothesis 3</i> .....	68
<i>Hypothesis 4</i> .....	79
<i>Hypothesis 5</i> .....	90
Conclusion .....	95
Chapter Five .....	96
Summary and Implications .....	96
Introduction .....	96
Overview .....	96
Discussion .....	97
<i>Research Question 1</i> .....	99
<i>Research Question 2</i> .....	102
<i>Research Question 3</i> .....	104
<i>Research Question 4 &amp; 5</i> .....	106
<i>Research Question 6</i> .....	108
Implications .....	110
Limitations.....	113
Recommendations for Further Research .....	115
Recommendations for Practice .....	118
Conclusion .....	120
References .....	121

## Tables

Table 1	
Research Questions and Corresponding Hypotheses .....	47
Table 2	
Descriptive Statistics: Hypothesis 1 (Overall Measurement) .....	51
Table 3	
ANCOVA: Hypothesis 1 (Overall Measurement) .....	52
Table 4	
Descriptive Statistics: Hypothesis 2a (Steps to Empowerment) .....	53
Table 5	
ANCOVA: Hypothesis 2a (Steps to Empowerment) .....	54
Table 6	
Descriptive Statistics: Hypothesis 2b (Credibility) .....	55
Table 7	
ANCOVA: Hypothesis 2b (Credibility) .....	55
Table 8	
Descriptive Statistics: Hypothesis 2c (Challenge and Support the Organization) .....	56
Table 9	
ANCOVA: Hypothesis 2c (Challenge and Support the Organization) .....	57
Table 10	
Descriptive Statistics: Hypothesis 2d (Build Teams) .....	57
Table 11	
ANCOVA: Hypothesis 2e Hypothesis 2d (Build Teams) .....	58
Table 12	
Descriptive Statistics: Hypothesis 2e (Develop Individuals) .....	59
Table 13	
ANCOVA: Hypothesis 2e (Develop Individuals) .....	59
Table 14	
Descriptive Statistics: Hypothesis 2f (Quality and Task) .....	60
Table 15	
ANCOVA: Hypothesis 2f (Quality and Task) .....	61

Table 16	
Descriptive Statistics: Hypothesis 2g (Customer/Client) .....	61
Table 17	
ANCOVA: Hypothesis 2g (Customer/Client) .....	62
Table 18	
Descriptive Statistics: Hypothesis 2h (Vision) .....	62
Table 19	
ANCOVA: Hypothesis 2h (Vision) .....	63
Table 20	
Descriptive Statistics: Hypothesis 2i (Foster Creativity) .....	64
Table 21	
ANCOVA: Hypothesis 2i (Foster Creativity) .....	64
Table 22	
Descriptive Statistics: Hypothesis 2j (Manage Diversity) .....	65
Table 23	
ANCOVA: Hypothesis 2j (Manage Diversity) .....	66
Table 24	
Descriptive Statistics: Hypothesis 2k (Build Public Trust) .....	67
Table 25	
ANCOVA: Hypothesis 2k (Build Public Trust) .....	67
Table 26	
Descriptive Statistics: Hypothesis 3a (Steps to Empowerment) .....	69
Table 27	
Descriptive Statistics: Hypothesis 3b (Credibility) .....	70
Table 28	
Descriptive Statistics: Hypothesis 3c (Challenge and Support the Organization) .....	71
Table 29	
Descriptive Statistics: Hypothesis 3d (Build Teams) .....	72
Table 30	
Descriptive Statistics: Hypothesis 3e (Develop Individuals) .....	73

Table 31	
Descriptive Statistics: Hypothesis 3f (Quality and Task) .....	74
Table 32	
Descriptive Statistics: Hypothesis 3g (Customer/Client) .....	74
Table 33	
ANCOVA: Hypothesis 3g (Customer/Client) .....	75
Table 34	
Descriptive Statistics: Hypothesis 3h (Vision) .....	76
Table 35	
Descriptive Statistics: Hypothesis 3i (Foster Creativity) .....	76
Table 36	
Descriptive Statistics: Hypothesis 3j (Manage Diversity) .....	77
Table 37	
Descriptive Statistics: Hypothesis 3k (Build Public Trust) .....	78
Table 38	
ANCOVA: Hypothesis 3 (Overall Measurement) .....	79
Table 39	
Descriptive Statistics: Hypothesis 4a (Steps to Empowerment) .....	81
Table 40	
Descriptive Statistics: Hypothesis 4b (Credibility) .....	82
Table 41	
Descriptive Statistics: Hypothesis 4c (Challenge and Support the Organization) .....	82
Table 42	
Descriptive Statistics: Hypothesis 4d (Build Teams) .....	83
Table 43	
Descriptive Statistics: Hypothesis 4e (Develop Individuals) .....	84
Table 44	
Descriptive Statistics: Hypothesis 4f (Quality and Task) .....	84
Table 45	
Descriptive Statistics: Hypothesis 4g (Customer/Client) .....	85



Table 46	
Descriptive Statistics: Hypothesis 4h (Vision) .....	86
Table 47	
Descriptive Statistics: Hypothesis 4i (Foster Creativity) .....	87
Table 48	
Descriptive Statistics: Hypothesis 4j (Manage Diversity) .....	88
Table 49	
Descriptive Statistics: Hypothesis 4k (Build Public Trust) .....	89
Table 50	
ANCOVA: Hypothesis 4 (Overall Measurement) .....	90
Table 51	
Descriptive Statistics: Hypothesis 5 (Overall Measurement) .....	92
Table 52	
ANCOVA: Hypothesis 5 (Overall Measurement) .....	93
Table 53	
Numeric Differences in Means .....	94
Appendices	
Appendix A	
Leadership and Commitments and Credibility Inventory Self Form .....	130
Appendix B	
Leadership and Commitments and Credibility Inventory Other Form .....	141
Appendix C	
Validity and Reliability Analysis – University of Connecticut .....	154
Appendix D	
Leadership and Management Science Unit Letter to Course Attendee .....	174
Appendix E	
Technical Appendix: Reliability Analyses - Time One .....	178
Appendix F	
ANCOVA Results Hypothesis 3 .....	180
Appendix G	
ANCOVA Results Hypothesis 4 .....	186

Appendix H	
ANCOVA Results Hypothesis 5 .....	192
Vitae .....	198

## CHAPTER 1

### Introduction

#### Overview

As the 20<sup>th</sup> Century ended, the use of systems instruments procedures became popular tools for the improvement of leadership practice in organizations. A subset of these systems instruments is multi-rater feedback, known also as 360-degree or multi-source feedback (Warech, Smither, Reilly, Millsap, & Reilly, 1998). These systems are generally designed for the improvement of organizational effectiveness and involve various components. Specifically, the multi-rater feedback process involves systematic assessment from subordinates, peers, supervisors, and customers, all done to solicit perceptions about an individual's practices. This approach assumes that performance varies across contexts, and that leaders will act differently with dissimilar constituencies. By receiving feedback from more than one layer within a bureaucracy, each with whom the individual being rated regularly interacts, the executive's overall understanding of how they affect others will be enhanced.

The positive aspect of multi-rater feedback is that it directly acknowledges differences in viewpoints from various opportunities of observing a leader's behavior (Day, 2000). Typically, the information is used for developmental purposes, although it has been tied to performance evaluation (though not as used for this study), and can be a powerful tool for leadership improvement. It has become a widely accepted piece of human resource development programs, and is assumed to have value in the work environment context.

According to Dalton (1998) "it is difficult to find a Fortune 1000 company in the United States that has not tried a 360-degree assessment somewhere in the organization at least once" (p. 59). Additionally, Atwater and Waldman (1998) estimate 90 percent of Fortune 1000 firms use some form of multi-source feedback. As we move into the 21<sup>st</sup> Century it appears as if this technique will become a more widely accepted mechanism for the development of leaders.

As the popularity of this type of assessment has become more extensively acknowledged, it has carried over to the public sector. Additionally, as taxpayer funded

organizations move forward, they like their private sector counterparts, are likely to constantly seek opportunities for improved performance. Multi-source evaluation of their leaders is believed to be one of these methods.

As Loehr (1997) describes in a discussion about feedback and corporate executives, “you’ve got to open up the hood and take an honest look at what’s really underneath it. No hiding. No denying” (p. 36). He continues, “opening up to personal truth can be both scary and painful, but the payoff in growth can be tremendous” (p. 36).

And why shouldn’t it be? If our corporate and public sector leaders take an honest and critical look at themselves on a consistent and on-going basis, not only will their leadership practice improve, but those who model themselves after the leader will also see growth in their leadership abilities. Moreover, respondents who fill out a multi-rater evaluation of their leaders receive a reinforcing message of the importance of the items, which are part of the assessment. That is to say, many hundreds of individuals within an organization will complete the multi-rater assessment for the recipients of the feedback, and in that way be reinforced on what is the value message for the organization. “Thus, it contributes to individual, group, and organizational development...and sensitizes employees to the organization’s expectations” (London & Tornow, 1998, p. 4).

#### Conflicting Opinions on the Use of Multi-rater Feedback

There are conflicting views about the best use of multi-rater feedback (Bracken, Dalton, Jako, McCauley, & Pollman, 1997). The debate surrounds the confidentiality of the information and exactly how that information might be used. According to Dalton (1998), “using 360-degree feedback for appraisal has two implications for the process: (a) the immediate boss will see the results; they will not be confidential, and (b) the individual will be rewarded or punished (salary increases, promotions, and so on) based on the results” (p.69). However, Bracken et al. (1997) argue that providing feedback without using other aspects of an organization’s operational and administrative process denies the assessed individual the full benefits of the feedback procedure.

While this debate is ongoing, there continues to be a lack of complete insight into all the subtle nuances of the use and effects of multi-rater feedback, and it has become a

research domain ripe for study. “Unfortunately, what we know about the integration of the large amount of data generated by 360-degree feedback is limited” (Brutus, Fleenor, & London, 1998, p.27).

The present research provides some insight into the nuances of multi-rater feedback by examining the area of perceptual agreement among multi-rater feedback sources. It adds to the literature by offering insight into the use of multi-rater assessment as part of a case study in a public sector law enforcement organization – the Federal Bureau of Investigation (FBI).

### 360-Degree Feedback and Development

London et al. (1998) suggest that improvement in the leadership domain can be accomplished through the use of the 360-degree feedback process. Edwards and Ewen (1996) indicate that individuals and organizations using multi-rater feedback can gain a better understanding of their organizational strengths and weaknesses, assess the quality of their leadership as compared to norms, and forecast their training needs. According to Goldsmith (1996, unpaginated), “senior management teams at some of the world’s leading companies – American Express, Avon, GE, Netscape, Nortel, Texaco – use 360-degree feedback as part of an overall process to help align corporate values and individual behavior.”

According to work done by Hazucha, Hezlett, and Schneider (1993), multi-rater feedback can improve skill and development efforts. They found effectiveness was improved two years after the provision of multi-rater feedback.

Three hundred sixty-degree feedback can also be used to initiate developmental plans for improved performance. According to Nelson (1993), individuals are driven by a self-regulation process, which is routinely tied to a goal or benchmark. When individuals see that their performance does not match the standard of behavior that they originally attributed to themselves (as measured by the feedback), they adjust their behavior accordingly. Unfortunately, little is known about how managers receiving multi-rater feedback set goals to change their behavior (Brutus et al., 1998).

While the present research did not address goal-setting in conjunction with multi-source assessment (as this is another area suitable for further examination), it did focus

on perceptual agreement between multi-rater feedback sources. Specifically, it examined the agreement between perceptions of self and others in dimensions measured by the survey instrument used for this study. Additionally, it separated “others” into peers and subordinates, and examined differences among these assessments and self-assessments. By doing this, this research helped clarify the importance of subordinate and peer ratings in multi-rater assessments. This was done through the examination of hypotheses that sought to determine how perceptual agreement might change over time. To do so, a system designed to measure leadership dimensions considered important in law enforcement, particularly the FBI, was used.

#### Multi-rater Instrument Used for this Study

The process used to gather data for this study consisted of a multi-rater system in use in the FBI. This system is composed of two data collection instruments, called the “Leadership Commitments and Credibility Inventory System” (LCCIS). The primary instruments of this system are completed by subordinates (direct reports) of those being evaluated, peers, and by the individuals (self-reports) who are the focus of the feedback themselves. These instruments were created by Keilty, Goldsmith, and Company and the FBI and are attached as Appendices A and B. These instruments, their attendant characteristics, and their prior validity and reliability are discussed in Chapter 3.

#### The Federal Bureau of Investigation

The FBI has focused on leadership since the organization was founded in 1908. From its beginning, the FBI has constantly refined its leadership strategy in the law enforcement arena. When J. Edgar Hoover became its director, he was determined to establish the FBI as the preeminent law enforcement agency in the United States. Over time, and as a result of many successes, the FBI grew into its role as the leading law enforcement agency in the world (DeLoach, 1995). Today, the FBI continues to enjoy a worldwide reputation for leadership excellence in law enforcement.

While the FBI enjoys an enduring reputation, opportunities for improvement are evidenced by such occurrences as the widely publicized problems associated with the Ruby Ridge Incident, the siege at Waco, Texas, and arrests of FBI agents accused of

espionage. These episodes, as well as others for which the FBI has been criticized, all tend to suggest that leadership, or lack thereof, was a central factor of the final outcome. And while it is easy to focus on the negatives (for they receive media attention and capture the imagination of the American public), the FBI also routinely handles high profile cases that are executed flawlessly. This is due in no small part to the quality of leadership demonstrated within the organization.

Today the FBI is involved in a number of leadership initiatives, including training and development programs. The FBI Academy, located on the Marine Corps Base in Quantico, Virginia is on the cutting edge of these issues in the law enforcement profession. One such curriculum is the FBI's National Academy program where midlevel and executive police officers from agencies all over the world are invited on a regular and ongoing basis for training. National Academy participants reside at the FBI Academy for 10 or 11 weeks and receive instruction ranging from forensic science to leadership and management.

Many other training programs for law enforcement personnel occur at the FBI Academy, including training for future leaders of the FBI. Two of these programs are the Executive Development Institute (EDI) I for first-line supervisors and the EDI II for mid-level leaders. In both of these programs, management issues are introduced and professionals considered to be the best in their fields are invited to facilitate discussions regarding leadership. Additionally, as part of an effort to raise the level of leadership skill sets at the FBI, in 2001 the administration of the multi-source feedback instrument was approved for use in each of these programs.

Training for the FBI's administrators is seen as extremely important as it is the FBI's belief its executives should model the behavior it wishes other law enforcement agencies to adopt as their standard. In fact, part of the FBI's mission statement articulates its responsibility to provide assistance and leadership to local law enforcement. The statement reads, "The Mission of the FBI is to uphold the law through the investigation of violations of federal criminal law; to protect the United States from foreign intelligence and terrorist activities; to provide leadership and law enforcement assistance to federal, state, local, and international agencies; and to perform these responsibilities in a manner that is responsive to the needs of the public and is faithful to

the Constitution of the United States” (Federal Bureau of Investigation, n.d.). Clearly, modeling leadership and providing training for personal and professional development in the leadership domain is a matter the FBI takes very seriously.

#### Adult Learning and Human Resource Development

The concept of providing training for adults focusing on personal and professional development is a cornerstone of adult learning and human resource development (ALHRD). ALHRD incorporates research that focuses on social change through the use of leadership development and group dynamics. Additionally, ALHRD focuses on personal growth and organizational development (Stubblefield & Keane, 1994). As an extension of this process, many companies have adopted leadership development programs for their executives. The FBI’s leadership training has been designed in this vein, specifically, for leadership improvement and skill enhancement.

The present research focused on a unique part of the executive growth and enhancement process, and sought to examine aspects of leadership and development in the FBI through using a multi-source feedback system. This point is significant because of its particular focus on leadership in the executive ranks of law enforcement. Specifically, the study examined perceptual agreement between multi-rater feedback sources of FBI executives at two points in time. Only two time points were used because this study was constrained by a finite research period.

#### Leadership

A variety of researchers have examined successful leaders and organizations to determine what constitutes success (Bennis & Nanus, 1985; Kouzes & Posner, 1993; Peters & Waterman, 1982). Furthermore, many studies have been published to support trait theories of leadership (Yukl & Van Fleet, 1992). In fact, in the United States trait theories of leadership were the conventional wisdom until the 1940s, when a strong belief arose that leaders were born and not made. This approach highlighted the attributes of individuals “such as personality, motives, values, and skills” (Yukl, 2002).

During the latter half of the 20<sup>th</sup> Century a change occurred and ideas about leadership and management were narrowed to leadership development and the growth of



the individual executive. As the 1990s approached, the focus shifted to the full development of the organizational leader. Aspiring leaders were required to inspire and communicate a shared vision, demonstrate empathy, improve leadership and management skills, model the way, and have knowledge of oneself (Bennis & Nanus, 1985; Kouzes & Posner, 1987; Nadler & Tushman, 1990).

Regarding the responsibilities of leadership in the 1990s, Spencer & Spencer (1993) suggested that distinguished superior leaders recognize their own limitations by unambiguously affirming them and work on triumphing over them. Schein (1992) offered that leaders need to have a high level of objectivity about themselves and their organizations, which requires them to see and understand their growth areas. This includes recognizing and managing their anxieties and defensiveness. Staub (1996) boldly stated “we need the courage to challenge the old assumptions and to ask tough questions about our ways of interacting, thinking, and living. Then we need the courage to take in new data and act on it” (p. xvii).

These authors express daunting challenges for leadership in both the private and public sectors, and the difficulties will be no less significant in the future. Therefore, it is imperative that present-day leaders maximize opportunities to learn about both their strengths and growth areas. Through the use of multi-source feedback, individuals should be able to see themselves through the perceptions of others and use these insights to improve the way they lead in their organizations.

## The Issues

Do supervisors in a work environment change their perception of their leadership ability after having the opportunity to examine how others see them in this role? Over time, and after reflection on their own and others’ perceptions, do these individuals change the way they see their leadership skills? Does multi-rater feedback on supervisors’ leadership dimensions help supervisors see themselves the way their subordinates and peers do in a public sector setting? Is this effect particularly salient in a law enforcement one? Is multi-rater feedback effective, and is convergence of perceptions between self-perceptions and others’ perceptions desirable?

This study considered the ideas expressed in these questions. Specifically, elements surrounding these questions were examined as a case study in a public sector law enforcement setting, the FBI.

Research in this domain is notable in that the use of multi-source feedback is now being done with regularity in the private sector, but is still being used in an even more limited way in the public sector, particularly law enforcement. The present research therefore adds to the literature. It focuses on using multi-rater feedback as a basis for examining how perceptual agreement may change over time.

Furthermore, the present study made the effort to identify whether or not leaders in the FBI have greater perceptual agreement between others' assessments and their own assessments on subsequent evaluations. It examined differences in self-assessments versus others, measured by the same instrument at two points in time. It also hypothesized that these assessments at time two would be in greater agreement than at time one. Underlying this was that after the initial feedback, leaders should have a more accurate perception of their impact on others, and on second assessments self and others' ratings should converge even more.

According to Hazucha et al. (1993), individuals who have been assessed by 360-degree feedback on more than one occasion increased their leadership skill level on second assessments. This research partially tested Hazucha's et al. (1993) contention. It assessed whether or not individuals who were evaluated with a similar mechanism on more than one occasion increased their leadership on second assessments. It did so by subdividing categories into leadership sub-scales and determining if differences that were found between self-perception and others at time one converged at time two. To add further depth, this study also examined whether there were differences in agreement between peer and subordinate ratings at time one and time two.

## Research Questions

This study examined the following questions. Data for this research were collected from individuals who participated in two multi-rater assessments at two different points in time. The time span between evaluations varied but was generally three years.

1. Is there a greater perceptual agreement between others (peers and subordinates) and self on the second multi-rater assessment than the initial assessment?
2. What are the dimensions (sub-categories of the LCCIS) of any differences found?
3. Are there differences in agreement between peer and self-ratings and subordinate and self-ratings at time two and time one?
4. What are the dimensions of any differences found in peer and self-ratings?
5. What are the dimensions of any differences found in subordinate and self ratings?
6. Are there differences in agreement between peer and subordinate ratings at time one and time two?

#### Definition of Terms

The term “multi-rater feedback” is defined for the purpose of this research, as:

*Multi-rater feedback* – Multi-rater feedback concerns receiving information about one’s leadership abilities from a variety of different sources, including subordinates and peers, in the form of written documentation. Multi-rater feedback as a concept can be further broken down into the terms “270-degree feedback” and “360-degree feedback.” 270-degree feedback is multi-rater feedback from the assessed individual, subordinates, and peers, and is the system in use by the FBI. 360-degree feedback is the process of receiving information about oneself from self, subordinates, and peers. It adds the layer and perspective of the individual’s supervisor. For purposes of this research, multi-rater feedback and 360-degree feedback will be used interchangeably.

The term “subordinate” is defined for the purpose of this research, as:

*Subordinate* – Someone who works for and in no case is higher in grade level or position than the assessed individual.

The term “peer” is defined for the purpose of this research, as:

*Peer* – Someone who is reasonably close or equal to the assessed individual in both grade level and or position.

The term “self-report” is defined for the purpose of this research, as:

*Self-report* – Someone who is a mid level executive in the organization who has supervisory responsibilities over people and programs. There is little or no difference in variation among those who were assessed in this study in terms of both grade level and position.

#### Outline of the Remainder of the Study

Chapter 2 relates and reviews literature surrounding the research areas: perception, multi-rater feedback, and leadership. Additionally, it examines the area of perceptual agreement between multi-rater feedback sources.

Chapter 3 describes the context, setting, participants, raters, instrument, and research procedure. It concludes with the data analysis plan.

Chapter 4 analyzes the data that address the research questions. It focuses on the phenomena of interest, including the findings.

Chapter 5 provides the discussion, summary and implications of this research. Finally, it describes the limitations of this research, and concludes with recommendations for further study and practice.

## CHAPTER 2

### Review of Literature

#### Overview

The present study sought to identify whether or not there was perceptual agreement in multi-rater feedback sources in the FBI, as indicated by dimensions of leadership measured at two points in time. Therefore, a review of the literature related to perception lends perspective to this study. Additionally, a literature review of multi-rater feedback provides insight into the specific focus of the research, and a review of that written about leadership provides further depth. Finally, to offer particular focus to the contribution of this research, a review of the literature related to perceptual agreement between multi-rater feedback sources is provided.

#### Foundation

Multi-rater feedback has its roots in behavioral science research, social psychology, differential psychology, and individual and organizational development. Writing in 1993, Dunnette stated, “the 360-degree feedback movement draws from and depends upon an exceptionally wide range of topics that have been represented for years in our behavioral science journals – such topics as *person perception, impression formation, individuals’ conceptions of self, impression management, and behavioral change*” (p. 374). Yammarino and Atwater (1993) reviewed these areas offering that in the literature there seems to be an ongoing and never-ending effort designed to understand self-perception, person perception, and behavioral change efforts.

Because this research is examining perceptual agreement as it relates to examining oneself and others, it is important to explore what the literature suggests are some of the significant aspects of the perceptual process in the domain of human interaction and social intercourse. Therefore, this literature search begins with the concept of perception.

## Perception

“The process of perceiving other people is rarely translated (to ourselves or others) into cold, objective terms...More often, we try to get inside the other person to pinpoint his or her attitudes, emotions, motivations, abilities, ideas, perceptions, and traits” (Knapp, 1978, p. 88). Knapp elaborates that each of us are a product of our cultural, educational, and personal backgrounds, and that our perceptions are influenced by all of these conditioning experiences. Additionally, he suggests that we choose to see certain things about others, and what we see is often dependent upon how we process this information.

As early as the 1940s, Asch (1946) suggested we form opinions of others based on an overall perception. Once this is done we slant assessments of individual traits to the perception we have taken as a whole of the individual. Specifically, if we see another person as offensive, anything associated with this person’s particular idiosyncrasies will also be perceived as negative.

First impressions play a role in our perceptions of others. Our visual cues cause us to draw immediate direct inferences, which we will often then try to validate over time. Knapp (1978) explains that first impressions are used to reduce uncertainties and to avoid having to admit mistakes when judging others. His theory suggests that we avoid changing our opinion even in the face of contradictory information, in order to cling to an original view. While assessments can change over time, initial perceptions can be very powerful and are often lasting. This has also been called primacy effects.

Perception determines the quality and quantity of our exchange of ideas (Laborde, 1997). She suggests that we perceive the world with our senses and code the information on a map in our brain, which then determines our behaviors. Our behavior includes how we describe and interact with others. Furthermore, she suggests that most of our perceptions are unconscious and that our memories are selected pictures, sounds, and feelings, which we have condensed to fit our belief system. She also suggests that this process works in reverse and that, instead, our belief system helps us select the pictures, sounds, and feelings we hold on to. This includes how we see others and ourselves.

### *Self-Perception Versus Person Perception*

In the 1990s Yammarino and Atwater (1993) indicated there was an ongoing effort designed to understand self-perception and person perception in the literature. Both person perception and self-perception include insight, awareness, acuity, discernment, observation, and sensitivity, which are all aspects of the multi-rater feedback evaluation process. Self-perception can be described as a subset of person perception. Both person perception and self-perception are discussed below.

According to Schneider, Hastorf, and Ellsworth (1979), when an individual makes a perception about another person they make their judgment based on the person, the behavior, and the context. This process is known as person perception. Person perception is very complex, but usually begins when we observe or hear another person doing something. It includes the process of first impressions, judgment making, and assessing. The final stage of the person perception process usually involves making a prediction about the future behavior of the individual.

Self-perception has many overlapping characteristics with perceptions of others. In research conducted by Locke (2002), he concludes that despite common beliefs, individuals think the same about themselves and others in terms of complexity. He suggests this is important because this finding may influence social cognitive theory and research, and “differences in the complexity of how people conceptualize the self versus others might predict differences in how people respond to and evaluate the self versus others” (Locke, 2002, p. 1094). Further, according to Schneider et al. (1979), insights about self and others naturally reinforce one another.

However, Jones and Nisbett (1972) found that the differences between self-perception and others’ perceptions tend to be that individuals attribute their behavior to situational circumstances, while observers tend to see others’ behaviors in terms of internal dispositional explanations. A common assumption is that individuals have greater episodic knowledge of the self than others (Locke, 2002). This aspect of perception has relevance in the present study, as it was hypothesized that individuals would rate themselves more highly than their peers and subordinates on their initial multi-rater assessments. Supervisors may see their reasons for doing things as

situational, whereas others may see them as permanent aspects of their personality and character, including their leadership credibility.

To add further depth to the concepts of episodic and situational assessments of self, Markus and Wurf (1987) suggest that when an individual is thinking about and describing him or herself they do so with knowledge they have in their short-term memory. They use attributes they have in their working self-memory to describe themselves. However, that type of memory holds only a very restricted amount of information. This may also be true when individuals assess others. These aspects of the person perception process have implications for the present research, although they are not being measured. This is another fertile area for study.

Another aspect of person perception and self-perception that has potential relevance for the present study, is the research done by Brown (1986) who suggests that individuals tend to describe themselves more positively and less negatively than they do others, though not more positively than descriptions of a best friend. If assessing a best friend or someone for whom there is a positive general impression, there is potential for a “halo effect.” This idea suggests that if the overall impression is positive, the assessed may be rated highly in all areas even though some of their leadership traits are flawed.

The present research asks respondents of the feedback to have someone else distribute the surveys, and, therefore attempts to minimize the potential for the halo effect.<sup>1</sup> This research took the position that by having a neutral party hand out the instruments there was potential for a more objective evaluation than if the assessment tools were given out by the assessed to individuals of his/her choice. Unfortunately, this type of an intangible effect is hard to completely account for in social science research and must be considered as a potential limitation on the results of the present study.

#### *Other Biases and Problems in the Perception Process*

Certain other biases were discovered in the literature relating to the perception process. These include contrast effects, leniency, central tendency biases, and recency effects (Kreitner & Kinicki, 2003).

Contrast effects deal with the perceiver’s perception of the individual on whom the judgment is being made. If the assessed individual is among low performers he/she

---

<sup>1</sup> See Appendix D for instructions regarding distribution of survey instruments.



may be rated more highly than if that person is among high achievers. Leniency and central tendency biases deal with how a rater normally assesses others. Some raters will always be overly lenient regardless of the person being rated; some will always avoid extreme judgments, etc. Finally, there is recency bias. This perception quandary deals with recent information. Regardless of whether it is positive or negative the person is rated as such.

### *Perception Change Theories*

There is also literature that discusses theories about changes in perception as a result of a reflective process. This concept has some bearing on the present study, as it is hypothesized that reflection will bring about a convergence of rating scores between self and others on second multi-rater assessments as compared to first assessments. Therefore, the reflective process is directly associated with perception and this is reflected in the literature through perception change theories.

Some of the theories regarding perception change give rise to philosophical and ethical issues. These theories have broader cognitive focus than those ideas discussed earlier, in that they focus on the mental processing of experience and personal understanding. Mezirow (1990) offers a theory in this domain, which incorporates the concept of critical reflection. About his theory of perspective transformation he proposes, “Perspective transformation is the process of becoming critically aware of how and why our presuppositions have come to constrain the way we perceive, understand, and feel about our world; of reformulating these assumptions to permit a more inclusive, discriminating, permeable, and integrative perspective; and of making decisions or otherwise acting upon these new understandings. More inclusive, discriminating, permeable, integrative perspectives are superior perspectives that adults choose if they can because they are motivated to better understand the meaning of their experience” (p.14). This idea may be relevant to the present study, as it is believed that following the first iteration of the multi-rater assessment there will be reflection on the feedback, which will allow for a superior perspective of greater reality in the second assessment of self.

Finally, Freire (1970) espoused another perception theory more radical in its impact on the social conscience. His theory relates to how adult learning is played out through the education process. He suggests that in an educational encounter those who

are receiving the education not only gain a deepening awareness, but also a capacity to transform their lives. His theory centers on social change, particularly for oppressed people and, similar to Mezirow's theory, has philosophical and ethical considerations attached to it because of the motives and responsibility of those providing the education.

Theorists such as Mezirow and Freire offer greater conceptual frameworks than can be focused on in this research. However, as far as the present study is concerned, it is important to understand the multifaceted nature of the perception process and, by extension, the potential for change associated with it. Further, concepts that incorporate changes in perception can be sophisticated topics that define social movements and have been written about in this way in the literature. Mezirow and Freire offer theories that expose the seriousness of the responsibilities associated with the processes, whereby human resource development professionals expose individuals to information that may cause deep personal introspection and critical reflection about their perceptions. The present research examines perceptual introspection that may cause profound introspection in an organizational setting.

### Multi-rater Feedback

While the 360-degree feedback system has received much publicity and use, it is reportedly one of the least understood management developments in recent years (Church & Bracken, 1997). According to Jackson and Greller (1998) and Nowak, Hartley, and Bradley (1999), there is little definitive research concerning this process. Further, there is little information concerning multi-rater feedback and law enforcement in the literature.

One of the roots of 360-degree feedback is the employee attitude survey, which seeks to discern and lead to organizational change (Nadler, 1977). Additionally, London and Tornow (1998) offer that the source of multi-rater feedback can also be found in the performance evaluation process, used in most public and private sector organizations. They suggest 360-degree feedback is a significant learning tool and should be used as the core for self-development.

London and Smither (1995) write about a more practical foundation for the use of multi-rater feedback, suggesting it has grown in popularity out of the pressure for businesses to change in a more competitive marketplace. As evidence of this premise, in

1991 Van Velsor and Leslie suggested the previous 15 years had seen a dramatic growth in the number of such instruments. By 1996 Bohl found that 13 percent of companies surveyed used a 360-degree feedback instrument, which they designed and administered internally or the company purchased from another vendor, and another nine percent used some form of partial multi-rater assessment. The same year, Edwards and Ewen (1996) reported that 90 percent of Fortune 1000 companies used some form of multi-rater assessment. By 1998 Dalton offered that it was difficult to find a Fortune 1000 company in the United States that had not experimented with multi-rater feedback in one form or another.

#### *Process and Purpose*

Illuminating the concept of what 360-degree feedback involves, Hazucha et al. (1993) offer that multi-rater feedback instruments “provide managers with feedback about their skills from multiple sources: typically their supervisors, peers, subordinates, and themselves” (p. 325). London and Smither (1995) define the concept as the process by which performance evaluations of an employee are collected by more than one source. Brutus et al. (1998) succinctly state, “the ratings from self and others constitute the core of the 360-degree feedback process” (p.15), and the advantages of such a system are that each of the individuals, whether they are direct reports, peers, or the supervisor, may provide germane but slightly different information (Borman, 1991).

The 360-degree feedback procedure consists of three components: data, evaluation, and action (Jackson & Greller, 1998). The data are factual information obtained from raters who observe the person being measured. This data usually take the form of responses to questions, or narrative descriptions, which raters provide on the survey questionnaires. The evaluation element consists of a value judgment or reading of responses. The action element relates to what occurs as a result of the process. According to Jackson and Greller, the action element may be the key to effective utilization of a 360-degree feedback system.

These instruments are usually designed in such a way that the “capacities assessed in the tool are those that the organization wants its managers to develop over the long run” (Van Velsor, 1998, p. 150). As such, those filling out the instruments receive a reinforcing message of the importance of the items as they complete the forms. “Thus, it

contributes to individual, group, and organizational development...and sensitizes employees to the organization's expectations" (London & Tornow, 1998, p. 4). As Schein (1992) explains, groups "must develop a common language and category system that clearly define what things mean" (p. 92). Multi-rater systems are a means by which organizations can instill and strengthen a set of beliefs in their employees.

Studies of organizations show that the key to growth and survival is to keep the needs of their constituencies in some kind of balance. Defining the organization through a set of beliefs about its core competencies and basic functions in society reflects this balance (Kotter & Heskett, 1992). The process of communicating the organization's values through repeated use of the multi-rater feedback instrument is a very powerful way to buttress the organization's values.

#### *Differences from Traditional Feedback Process*

The multi-rater feedback process is unlike traditional performance evaluation, because it includes subordinate and peer ratings. Their addition can offer some advantages as each data source presents different perspectives (Smither, Wohlers, & London, 1995). "Thus feedback from more than one source should have more informational value than feedback from the supervisor alone" (Smither et al., 1995, p.62).

"Top down" ratings are typically associated with organizations through the performance evaluation process (Murphy & Cleveland, 1995) where individuals know clearly who is providing them with the feedback. However, with multi-rater feedback, the sources of the critique are often anonymous and, therefore, the information is more readily accepted by its recipients (Dalton & Hollenbeck, 1996). Additionally, London and Smither (1995) suggest that this process provides a richness and depth of insight to the evaluated individual, which they would not otherwise have available to them.

#### *Reasons Organizations Use Multi-rater Feedback*

London (1997) suggests that individuals are not always good at providing and receiving feedback, and, therefore organizations take the initiative to develop systems with the goal of improving relationships. Research has shown that multi-rater assessment can supplement and improve the communications process (Bernardin & Beatty, 1987).

Wanguri (1995) lists some of the benefits intrinsic to multi-rater systems. Significantly, the multi-rater feedback system increases participation in the performance

management process. It is viewed as being more accurate than traditional feedback methods, and when used for promotional purposes, multi-rater assessments provide a better range and quality of information than “top down” systems. As a result, they tend to reduce or limit single rater bias.

Brutus et al. (1998) suggest that the impact of multi-rater feedback is extraordinarily powerful. This can be increased when an individual’s perception of self is different than the perceptions of others within the organization, as gained from the feedback process. Their premise is that the communication of the feedback is used by the organization to capture the feedback recipient’s attention.

London & Beatty (1993) state, “the 360-degree approach recognizes that little change can be expected without feedback, and that different constituencies are a source of rich and useful information to help managers guide their behavior” (p. 354). They offer that any appraisal of supervisory effectiveness must include managerial actions that are aligned with the organization’s values. A significant means of measuring this is through the multi-rater assessment system.

#### *Varying Uses of Multi-rater Feedback*

There are many ways that organizations use 360-degree feedback. In 1993 Tornow offered that multi-rater feedback is used as a means for organizations to change or improve performance levels, as a way to improve performance by increasing awareness or as a part of the performance evaluation process. London and Beatty (1993) offer that multi-rater feedback is the primary means for change and that through feedback individuals can experience powerful and lasting change. Tornow, London, and CCL Associates (1998) offer, “that 360-degree feedback should be a core part of self-development; that it fits the realities of the new workplace; and that it is multi-faceted in purpose and value” (p. xiv).

Perhaps the most popular uses of multi-rater feedback are employee development and/or performance appraisal. Of the two, employee development is the most frequently cited (Atwater & Waldman, 1998; Jackson & Greller, 1998). Due to concerns about misuse, many organizations use multi-rater feedback exclusively for development, and “only the managers being rated see the feedback” (Atwater & Waldman, 1998, p. 96). These two authors propose that the use of a multi-rater feedback system for

developmental purposes will usually produce more reliable results as compared to its use in an employee appraisal system. This is because the employee being rated, for the purposes of development, will be less likely to alter their behavior in order to influence rater scores. Further, by “playing to the raters,” those rated may use their influence and make decisions that may not be in the best interest of the organization.

In spite of the fact that multi-rater feedback is often given only to the recipient, it has been used in other ways as well. “In many cases, employees are given such feedback as part of a training or expanded appraisal program” (Smither et al., 1995, p. 61) or “within the context of a developmental program” (Sternbergh, 1998, p. 28). Sternbergh continues, “in some cases the feedback is integral to the program, that is, it supports the purpose of the program, and the program increases the value of 360-degree feedback. In other cases it is virtually an add-on: the feedback report is given to participants, but the program barely makes mention of it” (p. 28).

According to Dalton and Hollenbeck (1996), 360-degree feedback should be treated not as a one-time event but as a continuous process. Hazucha et al. (1993), argue that individuals who have been assessed with 360-degree feedback on more than one occasion increase their skill level on second assessments. Additionally, their work partially supports the premise that ability development will be related to both improvement efforts and environmental support, and that some developmental activities are more effective than others.

In 1994 Budman and Rice offered that the use of multi-rater feedback in the performance evaluation process was already becoming popular. In their examination in 1998, Atwater and Waldman specifically focused on whether or not an organization is ready to take multi-rater feedback beyond personal developmental use to become a part of the appraisal process. They suggested that in order for this to take place successfully, it must be gradually introduced into the culture. First, it should be introduced as a developmental tool until employees are comfortable with the process and later as part of the appraisal process. These authors acknowledged the debate associated with expanded use of the multi-rater feedback appraisal in the evaluation process, but suggest, “if handled properly, 360-degree feedback can broaden the perspective of performance appraisals and be a useful addition to the traditional performance process” (p.104).

London and Beatty (1993) offer a focused debate regarding the use of multi-rater feedback, particularly in connection with competitive advantage. In their work they present a set of measurement dimensions (including work unit and operational standards) that they suggest should be considered when designing and using a multi-rater feedback system.

#### *Multi-rater Feedback from Direct Reports, Peers and Self-Reports*

Feedback from subordinates is typically called “upward feedback,” and it involves the collection of ratings from multiple direct reports and the comparison of those ratings to self-ratings (Reilly, Smither, & Vasilopoulos, 1996). There are differences of opinion about how information from direct reports is treated. While feedback from direct reports may seem threatening, some studies have shown that feedback from subordinates is treated favorably (Bernardin, Dahmus, & Redmon, 1993).

There have been some interesting findings with regard to studies of upward feedback. For instance, a study by Wohlers, Hall, and London (1993) found that greater agreement between self and subordinate ratings arose when individuals were of the same race. Additionally, London and Wohlers (1991) found that direct reports’ assessments were more likely to be in agreement with female self-reporting than male self-reporting.<sup>2</sup>

With regard to critique from colleagues, Landy and Farr (1983) report that multi-rater feedback from peers was found to be valid and reliable. Brutus, Fleenor, and McCauley (1996) investigated demographic and personality predictors regarding self-reporting and multi-rater feedback. These authors concluded that such things as gender, race, self-esteem, introversion, age, and sensitivity influence self-perception, as well as the perceptions of others who are involved with the rating process.

#### *Goal Setting and Multi-rater Feedback*

According to London and Tornow (1998), “the 360-degree feedback process should include setting goals” (p. 1). However in many systems this is not usually done. For instance, London and Smither (1995) found that out of 20 organizations they studied, only eight consistently tied training measures to the results of the multi-rater feedback. Training measures can range from leadership development programs to a simple

---

<sup>2</sup> In the spirit of these two researchers, the present study considered examining the differences between genders with regard to perceptual agreement in the FBI. This was not done however after the data were cleaned and there was available information on only six female participants.

development plan. Of the remaining 12 companies studied by these two authors, only half tied the results to any type of goal setting. The rest did not have any procedure to ensure that those companies assessed set goals based on the feedback they received.

Another study done by Kluger and DeNisi (1996) examined the relationship between multi-rater feedback and goal setting. These two authors found that multi-rater feedback combined with goal setting results in superior return for the participants, compared with systems that administer multi-rater feedback alone. Sternbergh (1998) expands on this idea by offering that a good multi-rater system includes a program that concludes with goal setting.

Brutus et al. (1998) suggest that there is some certainty with regard to 360-degree feedback, and that is that multi-rater feedback provides the feedback recipient a range of information that can be used for goal setting. However, they also offer, “little is known about how managers receiving 360-degree feedback set goals to change their behavior and improve their performance” (p.27). This is another area suitable for further examination that was not addressed as a part of this research. It will be discussed in Chapter 5 under the “Recommendations for Further Research” section.

#### *Multi-rater Feedback and Law Enforcement*

Although there is little research regarding multi-rater feedback and the law enforcement community there are a few instances in the literature. Atwater, Waldman, Atwater and Cartier (2000) did a study regarding upward feedback using 110 supervisors from an unidentified policing agency. A unique comment offered in the limitations portion of their article provides insight into their thoughts, which may be reflected in the larger academic community regarding law enforcement. Specifically, they offer that the unique nature of the way police agencies are set up and the natural danger involved in this kind of work may have impacted their study. They continue, “police organizations are known for being quasi-militaristic and feedback from subordinates may have been perceived as an intrusion into the power structure” (p. 290). It is possible that these unique aspects of law enforcement have contributed to the lack of effort and/or need for the type of research being conducted here.

Currently, there are some examples of multi-rater and upward feedback in use by law enforcement. The Oceanside Police Department in California has been using



subordinate evaluations of their supervisors for at least five years. (See the Oceanside, California Police Department Subordinate Feedback Form.) The Fremont, California Police Department has a mandatory 360-degree feedback evaluation that is reviewed by the supervisor's evaluator and is attached to the supervisor's final evaluation. (See the Fremont, California Police Department 360-degree Feedback Form.) Additionally, the Arlington, Virginia Police Department has tested the use of subordinate ratings of their leaders in recent years. Finally, the Madison, Wisconsin Police Department has their officers evaluate their managers about the changes they need to make in order for their subordinates to perform more efficiently. This information is used for the overall benefit of the organization and for the self-improvement of the supervisor (Wycoff and Oettmeier, 1994).

Other policing agencies are also experimenting with the feedback process as they learn more about it through attendance at the FBI's National Academy Program. As evidenced by these examples, and the FBI's association with a multi-rater feedback assessment system for more than 15 years, it is clear that the law enforcement community is slowly experimenting with this form of adult learning and human resource development. Chapter 3 of the present study offers a discussion of the FBI's multi-rater feedback system and its idiosyncrasies, which is undoubtedly the most comprehensive one used in law enforcement and the foundation for this research.

## Leadership

Improving leadership capacity is one of the main purposes of human resource developmental systems. It is also a desired behavioral outcome (Brutus et al., 1998) of the effort behind the 360-degree feedback process. Leadership includes the power to influence. Creating the ability to influence others is achieved in a variety of ways, which include modeling the way, critically reflecting, nurturing relationships, achieving peak performance, and listening. The following literature search reviews these concepts, as well as others, and helps provide a foundation of understanding of how feedback can be a critical link to the improvement of leadership in organizations.

### *Importance of Leadership*

Leadership permeates all human endeavors. According to Maxwell (1998), “leadership ability is the lid that determines a person’s level of effectiveness” (p. 1). Additionally, he suggests, “everything rises and falls on leadership...[and] any endeavor you can undertake that involves other people will live or die depending on leadership” (p. 225). Drucker (1996) provides that successful leaders entrust others to do many things in their name, however, the most important things are those they do themselves.

### *Leadership as a Performing Art*

Kouzes and Posner (1995) state that, “leadership is also a performing art – a collection of practices and behaviors – not a position. Constituents don’t willingly follow positions; they follow people engaged in a process” (p. 30). These two authors also suggest that leaders model the way. They set an example through personal illustration and devoted execution. How they respond to and use feedback is a critical piece of modeling the way.

According to Staub (1996), “a leader knows how to read people – is always seeking to understand others with regard to their desires, dreams, strengths, and internal motivators, and is always seeking a more effective way of reaching out to them” (p. 98). Additionally, he proffers that this process is enhanced through the improvement of listening skills, including the feedback from others.

Goldsmith (1997, unpaginated) believes that as we forge ahead, organizations are moving from hierarchical structures toward “networked” associations. In part, this is being caused by technological changes as well as customer demands. Consequently, organizational leadership structures are changing. “In a hierarchical organization, leaders can more easily give orders and expect people to respond. In a network organization, leaders need to effectively influence people without line authority.” Therefore, developing one’s influence potential is a key piece of the leadership process.

### *Developing Leaders*

Maxwell (1993) submits that highly successful leaders develop not only followers, but also other leaders. Part of this process includes a succession planning system, which usually begins with talent identification and incorporates multiple assessment techniques such as simulations, written tests, interviews, and multi-rater

feedback. These kinds of assessments provide a risk-free way to identify leadership prospects, and offer insight into developmental needs.

Weber (1996) suggests that the future requires leaders to be role models. Additionally, he says, “we leaders of today, and the leaders in our organizations who follow us, must become skilled, committed growers of effective leaders, staffs, and governance people. We have no more important task” (p. 306).

Regarding leadership development, Tichy (1997) offers that executive employees need to have preparation in a concentrated and repeated form that asserts leadership growth at all levels in an organization. Business leaders need to embrace the concept that they can build other leaders throughout the organization and need to reward those who do so successfully. Effective leaders vigorously use their companies’ human resource development systems to cultivate leaders. In private industry, successful businesses and successful leaders understand that their responsibility is to take the most important asset they have, human resources, and make that asset even more valuable through personal growth and development.

#### *Leaders as Learners*

Donnithorne (1993) suggests, “Executive leadership is a state of constant learning. There can be no sitting back, no relaxing. Executive leaders are at the pinnacle of management but they will stay there only if they keep evolving. Indeed, leadership development is a process that never finishes” (p. 146).

Kaltman (2000) quotes Robert E. Lee regarding the importance of lifelong learning: “The education of a man or woman is never complete till they die. There is always before them much to learn and more to do. Our hardest lesson is self knowledge” (p. xiv).

Brookfield (1990) submits that learning takes place when there is a concerted effort to reflect on the experience. This process takes time and effort and should include experiences, which incorporate both an individual’s strengths and weaknesses. According to Bennis and Townsend (1995), a bona fide leader supplements their growth areas with others, and does the best they can with their inherent gifts. They constantly ask questions of themselves and acknowledge themselves for those things that they have done well. Leaders constantly are in search of ways to improve. Leaders also spend time

analyzing the good in order to produce more of it (Oakley & Krug, 1991). They focus on and analyze their entire existence in order to learn and grow. This is done through the critical reflection process.

### *Leadership and Reflection*

Bennis and Townsend (1995), suggest the best leaders are reflective practitioners who both think and act. These individuals understand how they affect other people. Additionally, they have a great deal of knowledge about themselves, including their strengths and weaknesses. They become leaders by developing an understanding of themselves by reflecting on experience. Leaders ask the right questions at the right time in order to discover the truth about themselves and their lives. Bennis and Townsend also suggest that good leaders should be good followers. “If you’re coming up within an organization, you must be a good follower or you’re not going to get very far. Leaders and followers share certain characteristics such as listening, collaborating, and working out competitive issues with peers” (p. 8). Receiving feedback and critically reflecting on it is the highest form of listening and self-discovery, and is one of the most significant ways to grow as a leader.

Covey (1989, 1991) lays out a foundation of personal reflection based around seven habits. He relates the process of self-discovery and leadership development through a story about sharpening a saw. In his story, a man is busy sawing a tree with a dull saw and is not making any progress. When asked why he does not take time to sharpen the saw, he replies that he is too busy. This parable vividly reflects an unexamined life. Sharpening the saw represents taking time to nourish the mind and soul in a way that allows for reflective thinking. Through this process an individual assesses what he/she is doing with their time and attention. Obviously, this is a very personal and soul-searching process, which relates to the leader’s personal and professional development, and is the essence of growth through the adult learning process.

In 1983 Clavell edited the work of Tzu, which offers that leaders need to be aware of themselves, their environment, and others and have a healthy understanding of where they stand in regard to their situation and those around them. Having this kind of knowledge is based on relationships and the willingness to be open to the suggestions of others. Only through continuous feedback does this kind of personal understanding

occur, which moves a leader toward peak performance, and separates those who are truly exceptional from the average performer.

### *Leadership and Feedback*

Staub (1996) proffers that leadership is about putting yourself in a position to be scrutinized by others. He says “being a leader means realizing that you influence and work with people in a 360-degree sphere: up, down, and around. It means that you are putting yourself on the line and inviting others to do the same” (p. 199). Further, he says that feedback allows learning to occur. Without feedback, individuals feel as if they are not being cared for, do not feel connected, and will not perform up to their ultimate potential.

Cohen (1990, 2000) believes that leaders are made and not born. He suggests that the process of how to become a good leader can be taught. As such, it is possible to grow in leadership capacity through the process of receiving and reflecting on feedback from others.

According to London and Tornow (1998), leadership can be improved through the use of a multi-rater feedback system. They suggest that it is possible to better understand the strengths and weaknesses of corporate leaders by using a 360-degree feedback system. This also helps to gauge how the quality of the leadership in their business measures up to other organizations.

De Vries (1994) suggests that leaders must have a vision and “need to have the knack of perceiving salient trends in the environment. They must be able to process many different kinds of information, and use their perceptions as a basis for judging the direction in which environmental forces are going” (p. 73). Without some kind of a feedback mechanism, this process is basically impossible.

### Perceptual Agreement Between Multi-rater Feedback Sources

The rationale for using multi-rater feedback to measure performance deals with the concept of self-perception (Atwater, Ostroff, Yammarino, & Fleenor, 1998). These authors suggest, “anonymous feedback from subordinates should help managers see themselves as others see them, and provide them with developmental feedback about needed changes in their behavior” (p. 578). This is important as research indicates that

self-perception as compared to others' perceptions seems to be an area where there is a lack of consistency (Harris & Schaubroeck, 1988). The present study sought to determine if these findings hold true among leaders in a law enforcement setting.

*Agreement Between Self-assessment and Others' Assessments*

One of the principal findings of multi-source research on the level of agreement between self-raters and other raters is that generally little agreement exists between self and others (Atwater & Yammarino, 1992). Furthermore, there is a propensity for individuals to overrate themselves as compared to others (Thornton, 1980; Harris & Schaubroeck, 1988).

Festinger (1954) suggests that individuals will discount the discrepancy between their concept of self and others when they are rated lower by others than they perceive themselves. This concept is called "cognitive dissonance theory." Bernstein and Lecomte (1979) offer that feedback, which is negative and not in alignment with an individual's self-perception, is not valued the same way as information in alignment with or higher than an individual's self-perception. Meyer (1980) suggested that defensiveness is associated with ratings that are different than self-ratings.

Atwater, Roush, and Fischthal (1995) offer that those who overrate themselves as compared to others tend to be inferior performers. Additionally, in research done by London and Wohlers in 1991, it was found that female managers tend rate themselves in line with direct reports' assessments more frequently than do their male counterparts. On the other hand, Shore and Thornton (1986) found no differences between gender in self-ratings versus superior ratings, although their study used non-managerial workers.

A study done by Halman and Fletcher (2000), where individuals rated themselves before and after a development center experience, found that pre-development center experience ratings displayed considerably less congruence between self and development center observer ratings than after the experience. Additionally, as a sub-component within their research, they found that under-raters became more accurate in their self-evaluations on post-development center assessments than did over-raters.

Atwater et al. (1998) offer, "assessments can be made of whether self-ratings are higher, lower, or in-agreement with ratings provided by others" (p.578). The present study examined the propensity for leaders to overrate themselves as compared to their

subordinates and peers on first versus second assessments, and whether or not there was a convergence of ratings on second assessments.

Pearce and Porter (1986) propose that feedback from others, which confirms an individual's self-assessment, is favored and viewed as more informative than feedback that does not confirm an individual's view of self. Pedler and Boydell (1980) suggest that feedback that is seen as exceedingly different than an individual's self perception will result in rationalization and will lead to an unwillingness to change or improve.

#### *Effect of Feedback on Self-Rating Perception Versus Others' Assessments*

There have been differences of research results of feedback on self-ratings versus others' ratings. For instance, according to Atwater et al. (2000), "the evidence regarding the impact of feedback on self-ratings is not altogether clear" (p. 278). This indicates some relevance for the present study. Ashford and Tsui (1991) highlight the advantage of individuals seeing themselves the way they are viewed by others. Additionally, Atwater and Yammarino (1997) offer that self and other agreement are related to effectiveness. Conversely, Fleenor, McCauley, and Brutus (1996) suggest that whether or not individuals rate themselves higher, lower or the same as others has no bearing on their job performance effectiveness. While this particular concept was not measured in the present research, it is another area open for study. Atwater et al. (1998) offer that "simultaneous consideration" of both self and others' assessments, no matter what their relationship, is of value for explaining performance outcomes.

Finally, Atwater et al.'s (2000) research indicates that leaders rating themselves and receiving feedback from others twice through multi-rater assessment lowered their self-ratings at time two and brought them more into alignment with their subordinates' ratings. They suggest that personal growth can take place if individuals who overrate themselves on a first assessment align their ratings with others on second assessments. The present research sought to examine whether or not Atwater et al.'s (2000) findings remain consistent within the context of the FBI work setting.

#### Summary

The literature reviewed here concentrated on perception, multi-rater feedback, leadership, and perceptual agreement between multi-rater feedback sources as a basis for

the present study investigation. This review offers particular focus to the contribution of this research. In the next chapter the methodology and procedures for the present study are detailed.



## CHAPTER 3

### Methodology and Procedures

#### Introduction

This chapter describes the research methodology and procedures used in the present study. The chapter begins with a brief overview of the research, and then describes the context of the study. It follows by addressing the setting, participants, raters, and the instrument. Finally, this chapter deals with the procedure and data analyses that were followed.

#### Overview

The present study sought to extend the research into multi-rater feedback. It did so by focusing on multi-rater assessments in a law enforcement setting. More acutely, it examined whether there is perceptual agreement between multi-rater feedback sources in the FBI.

#### Context

The public sector usually embraces changes in leadership and management developments more slowly than their private sector counterparts. However, the FBI has endeavored to stay abreast of the private sector. This evidences the fact that as a public sector and law enforcement leader, the FBI has consistently looked for opportunities to improve its leadership competencies. The incorporation of multi-rater feedback is no exception, as the FBI began using it to assess leaders more than 15 years ago.

The incorporation of a multi-rater instrument in the FBI was viewed by the institution as a novel idea at its inception and was not universally embraced. Initially, leaders within the organization were skeptical of its value and were fearful that the information would be used to harm the individual on whom the information was generated. Walter Sirene, who worked in the precursor to the Leadership Development Institute, was a very progressive leader and saw the opportunity for multi-rater feedback to add value to the development of leadership training in the FBI. Sirene was always studying the private sector to find opportunities which could be transferred to the FBI. In

his work with Keilty, Goldsmith and Company, he learned of their using this new management trend of the 1980's and embraced it for use in the FBI.

Sirene spent numerous hours convincing executives within the FBI that a multi-rater feedback system would benefit the organization. After many meetings with various leaders he was able to persuade one of the highest-ranking executives in the organization that multi-rater assessment was a good idea for personal leadership development. He also recommended that this individual be the first to be assessed with the system, as an example for others. The introduction of the "Leadership Commitments and Credibility Inventory System" (LCCIS) in this manner led the way for other executives to use the instruments.<sup>3</sup> It soon was instituted as part of a course called the Executive Development Institute (EDI) I, with the understanding that all results would be used only for personal development and that its outcomes would be shared exclusively with the individual on whom the information was generated.

The personal example of one of the FBI's highest ranking officials, coupled with the understanding that the information would be shared only with the individual on whom the information was generated, was the key to the organization's success in embracing the multi-rater feedback system. For more than 15 years, the process has been used for personal development with no evidence of abuse. The system was modified a few years after its incorporation into the FBI to capture a few more dimensions of leadership, which were not in the original version of the instrument. The instrument and its modification will be discussed later in this chapter.

Within the context of the FBI, this study took advantage of an opportunity to examine the results of a multi-rater feedback system. While this type of system is now being used with regularity in the private sector, it is being used to a limited extent in law enforcement. The present research adds to the literature by using multi-rater feedback as a basis for examining how perceptual agreement may change over time. It does so in the context of a system designed to measure leadership attributes considered important in law enforcement, particularly the FBI.

---

<sup>3</sup> The LCCIS is the multi-rater feedback instrument used for this study.

## Setting

The FBI is an organization that is known throughout the world. It built its reputation by being outstanding in investigative matters, centralizing fingerprints for use by all authorized police agencies, creating a world renowned laboratory, and bringing high standards to law enforcement (DeLoach, 1995). Additionally, due to its unique role of serving the public, leading in the law enforcement profession, training a myriad of United States and international law enforcement executives, and investigating the most complex national and international criminal conspiracies, the FBI has ongoing extraordinary responsibilities. Effectively carrying out these responsibilities requires trust. The public trust that the FBI enjoys today has developed over time and is a result of the fidelity of the agent and support personnel who serve the public in their role as FBI employees. In the work the FBI does, it is essential that this relationship built around trust not be breached.

However, it is noted that in the unsettled times of the last 20 years, greater candor has been demanded by the American public of taxpayer-funded institutions. As a result, the media and the public have raised a number of questions about the leadership of the FBI. Significant incidents such as Waco, Ruby Ridge, the arrests of FBI agents for espionage, the controversy surrounding the theft of nuclear secrets at Los Alamos, and the appearance of impropriety by one of the FBI's former directors, Judge William Sessions, have caused a search for ways to ensure the American public that high ethical behavior remains one of the organization's priorities. These incidents raise a question that probably cannot be answered, which is whether or not the use of the multi-rater system has helped to minimize the number of these types of incidents. And while the answer is difficult to determine, the FBI continues to examine ways to improve its leadership, including expanding the use of the multi-rater instrument.

Other steps the FBI has recently taken to deal with these and other issues of public trust and bring into focus the importance of leadership concerns was to rename the Management Science Unit at the FBI Academy, first the Leadership and Management Science Unit, then the Leadership Development Unit and finally, the Leadership Development Institute. While in part symbolic, these actions are representative of the

desire to change deep-rooted cultural beliefs about managerial behavior and emphasize the importance of leadership development.

The Leadership Development Institute is essentially a training, consulting, and program management entity responsible for teaching management and leadership theories and practices. The Institute maintains extensive liaison functions with local, state, federal and international law enforcement executives and provides leadership and consulting services to various units, divisions, and field offices of the FBI. In addition to many other duties and responsibilities, the Institute oversees the administration of the multi-rater instrument used in the FBI's executive programs, which is the foundation for this research. The present study is an extension of the work of the Institute.

Recently, as another step to improve its leadership programs, the FBI began evaluating its leaders with the same multi-rater instrument more than once during their career. These assessments occur approximately 18 months after their appointment as a supervisor, about three to five years later when they are promoted to middle management, and two to three years after that upon their promotion to a top executive leadership position. The present research examined two iterations of the use of the multi-rater system by the same individual. This was a practical decision, as it will take a number of years to develop enough cases for meaningful analysis of three iterations of the same person. It is believed that through multiple evaluations, those measured will use the results to change their behavior and improve their leadership performance.

The FBI has chosen to provide the information obtained from the multi-rater assessment only to the individuals being rated. None of the gathered information has been shared with anyone other than the individual being assessed, and it has only been used for personal developmental purposes. This is often the starting point when this type of assessment is being integrated into an organization (Atwater & Waldman, 1998). It is not the design of this study to offer an opinion about whether the use of information gathered with multi-rater assessments should be shared beyond personal use for the FBI, or any other organization – this is an area for further study. However, it was the purpose of this research to examine if a panel of those individuals, who already had the benefit of being assessed through this process, closed a gap between their self ratings and others' ratings on a subsequent assessment. And, while the results can be universally interpreted,

it is noted that the organizational context is likely to impinge on the leadership dimensions being noted by the instrument being used. This is a possible limitation to be discussed in Chapter 5.

Within the context of the FBI's use of this instrument, the sequencing of the different assessments used for the purposes of this study was as follows. The first assessment took place prior to a program called the EDI I. EDI I is a two-week leadership and management program, which is overseen by the FBI Academy's Leadership Development Institute. This course is a residential program held at the FBI Academy and the focus of the training is on personal and professional leadership development. The coursework includes instruction on creativity, political savvy, situational leadership, leading "Generation X," and communication skills for leaders.

Typically, leaders attending EDI I have been in a supervisory capacity for at least 18 months. Therefore, those evaluating them as part of the multi-rater assessment have had a very good opportunity to judge their strengths and growth areas. Because of the raters' fairly significant knowledge of those they are rating, the primacy effects should be minimized in the assessment process. Further research may determine the length of time it takes for first impressions to "wear away."

During EDI I, the results of the assessment are privately shared with the individual being rated and individual outcomes are assessed to examine strengths and areas where there is room for improvement. If they desire, course participants are given the opportunity meet with the multi-rater feedback contractor, who sits down privately and discusses procedures and strategies that may be useful for change in the areas of personal practice and professional growth. Strategies recommended can include reaffirming and building on areas of strength, setting personal goals, using resource guides to learn new methods and procedures, and assessing personal behavior patterns.

As part of EDI II, another residential leadership program for mid-level leaders in the FBI offered later in an executive's career, the same multi-rater instrument is now being used to rate attendees for a second time, and most of the individuals attending EDI II will have already been assessed with the multi-rater feedback instrument in their EDI I

experience.<sup>4</sup> This group is a relatively small one, since at the time of this study there were only three EDI II opportunities in which the multi-rater assessment was used. Those who had the second experience are the only ones being used for this study.<sup>5</sup> While it was not the focus of this research, an area for further study may be to compare leadership levels of those who have had the benefit of two versus one assessment. Such recommendations for further research will be offered in Chapter 5.

EDI II focuses on “operational” strategies and uses tabletop exercises, instruction by leaders in the FBI who offer examples of the organization’s best practices, and executives from FBI headquarters who have unique insights on how to interact with their component of the organization. Situational leadership techniques are discussed within the group setting, and emphasize those areas that are part of the multi-rater feedback instrument. Additionally, because many in the class were assessed with the feedback instrument a second time, results from this assessment were presented to the attendees with a repeated opportunity for one-on-one counseling with the contractor.<sup>6</sup>

According to Martineau (1998), the use of the same multi-rater feedback instrument on more than one occasion to examine the gradual development of competencies may be more valuable than using the instrument to measure the impact of a single intervention. Through this process an individual has the ability to determine if there is meaningful change over time, based on the two measurements used for comparisons. It is in this vein that this research was done. Additionally, Martineau argues that “instruments that measure multiple dimensions and provide feedback in a variety of formats are more likely to be of value when used multiple times than are less complex and dynamic instruments that give a single presentation of the results” (p. 234). “Value” in this context deals with worth to the individual being measured. The instrument used by the FBI measures a variety of constructs and offers multiple

---

<sup>4</sup> Because EDI I was not a mandatory program for first line supervisors until 2001, some of those attending EDI II did not have the benefit of the training offered during EDI I, nor were they assessed with the feedback instrument until their EDI II experience.

<sup>5</sup> A third and final opportunity for assessment is during a course for executive leaders called the Leadership Forum. This six-day program focuses specifically on leadership for senior executives and the multi-rater assessment is now being used for these participants as well.

<sup>6</sup> The contractor spent a half-day in this course with these more senior leaders offering personal counseling strategies, if asked for by course participants.

dimensions, which were evaluated and used in this study. These constructs and dimensions are articulated in the “Instrument” section below.

### Participants

The individuals who were selected for this study were those who had the benefit of being the recipients of more than one multi-rater assessment. At the time of this research, there were three EDI II programs since the new policy of using the multi-rater instrument more than once during an executive’s career had been instituted. Data on 73 individuals who were assessed with the instrument twice were gathered. These 73 cases represented the entire available pool of those with two multi-rater assessments through the end of the calendar year 2002, and were a random pattern of FBI personnel in their grade and position. Using the same 73 participants assured no systematic biases from substituting people. All data sources and information about individual participants were anonymous due to confidentiality agreements the contractor made with the FBI.<sup>7</sup>

### Raters

The raters in this study included subordinates, peers, and self-raters. In all three instances there was little variation within groups. Subordinates were reasonably close in the amount of time they had to observe each of the individuals being evaluated, as time in position of those assessed was relatively similar in all instances. Peers were equal in grade level and positional responsibility as those they were evaluating. Self-raters were also a homogeneous group as the in-service training they were attending (before which the assessments were administered) assured similar grade level and placement in the organization. Further, all self-raters were middle-management and aspiring executives in the FBI.

Raters at time one and time two were necessarily different individuals. Due to the transfer of supervisors to field office assignments throughout the United States with each promotional opportunity, it was very unlikely that instruments were disseminated to the same raters for any given individual during the two different assessments. This factor

---

<sup>7</sup> In part, these agreements stipulated that no results could knowingly be matched by name to any assessed individual and shared with a third party.

eliminated the potential for memory effects from first assessments to second assessments. Additionally, the amount of time between assessments (three to five years in most cases) eliminated awareness of artifacts for self-raters.

### Instrument

Each of the raters involved in this process were given a copy of the paper and pencil LCCIS to fill out confidentially. Only peers and subordinates were used as feedback providers. In this respect, this assessment was not a true 360-degree assessment, but a hybrid. In the present study it is being called a “270-degree assessment” as it omits an assessment from the individual’s immediate supervisor. Therefore, this assessment did not include feedback from the entire circle of relevant viewpoints. However, there was more than enough data for those assessed to view how others see them.

Keilty, Goldsmith, and Company developed the LCCIS for their private sector clients. The FBI initially used it in 1985 in its basic format. It then consisted of 77 questions grouped under eight categories including: 1) STEPS to EMPOWERMENT, 2) CREDIBILITY, 3) CHALLENGE and SUPPORT the ORGANIZATION, 4) BUILD TEAMS, 5) DEVELOP INDIVIDUALS, 6) QUALITY and TASK, 7) CUSTOMER/ CLIENT, and 8) VISION.

In 1991 the basic instrument was enhanced by the FBI to capture some of its unique dimensions of leadership responsibilities. As such, the contractor, with the guidance and counsel of individuals from the FBI’s Management Science Unit (the precursor to the Leadership Development Institute), modified the instrument to address issues considered critical to a public law enforcement organization. Three categories were added, including: 1) FOSTER CREATIVITY, 2) MANAGE DIVERSITY, and 3) BUILD PUBLIC TRUST. These three categories and the questions asked of respondents about them have the same appearance, formatting, and style as the questions in the rest of the inventory. The questions are grouped into a composite index measuring each concept. With the rest of the instrument, they were initially tested for validity and reliability.



The initial validity and reliability testing was done in November 2001 at the Leadership Research Institute at the University of Connecticut. The following analyses were conducted: confirmatory factor analysis, item analysis, and alpha internal consistency reliability analysis. Additionally, confirmatory factor analysis was used to examine the construct validity of the categories of the questionnaire. The results indicated the correlations of the items and categories were found to be quite acceptable. The category reliabilities for the questionnaire ranged from .85 to .95. Overall, the results of this initial validity and reliability testing showed the “instrument is psychometrically strong...” (Keilty, Goldsmith, & Co., 2001).<sup>8</sup>

At the contractor’s collection center, the information from respondents was entered into a database and a report was generated for each of the evaluated individuals. No information was shared with the FBI on any specific person. The individual’s report was confidential and was given only to that person on whom the information was generated. In this regard, this was truly a developmental exercise designed for self-improvement, and the information was not used in any way for direct performance evaluation or as information that might affect one’s chances for promotion.

The fundamental format of the instrument is a series of 97 questions grouped under different categories as follows:<sup>9</sup>

**STEPS TO EMPOWERMENT** – includes providing an enabling structure, supporting and collaborating, empowering others, and analyzing others’ readiness.

N = 20

**CREDIBILITY** – includes competence, character, composure, courage, and care for people.

N = 21

**CHALLENGE and SUPPORT the ORGANIZATION** – includes appropriately challenging decisions he or she thinks are not good for the organization and continuously striving to improve the organization.

N = 6

---

<sup>8</sup> The report describing the psychometric properties of the LCCIS is attached as Appendix C.

<sup>9</sup> N = # of questions under each category.

BUILD TEAMS – includes actively promoting teamwork and promoting other areas of the organization.

N = 4

DEVELOP INDIVIDUALS – includes providing developmental feedback on performance problems and correcting employee misconduct.

N = 7

QUALITY and TASK – includes maintaining high quality standards and explaining how the work results in organizational success.

N = 10

CUSTOMER/CLIENT – a customer or client is a person, internal or external, who receives a service or product. This construct includes following through on commitments made to customers and assessing customer satisfaction levels.

N = 6

VISION – includes explaining the organizational vision to others and implementing actions in the work group that are in alignment with the organizational vision.

N = 5

FOSTER CREATIVITY – includes encouraging others to creatively solve problems and allowing new ideas to be explored and evaluated.

N = 4

MANAGE DIVERSITY – includes providing equal opportunity to compete for assignments or promotions and dealing with complaints of discrimination effectively.

N = 5

BUILD PUBLIC TRUST – includes using position only for legitimate personal gain and integrity issues such as always upholding the law.

N = 9

Each of these categories contains questions that seek to get at the essence of the following query (that is taken directly from the instrument regarding the individual being evaluated): “While considering the individual’s effectiveness, how satisfied is the respondent with this person’s performance?” Responses are provided within the context of a Likert scale format. The Likert scale has a range of responses from “highly satisfied” to “highly dissatisfied.” Between these two extremes are the responses

“satisfied,” “neither satisfied nor dissatisfied,” and “dissatisfied.”<sup>10</sup> These responses are designed to measure the degree of perceived effectiveness of the leadership practices of the individual being evaluated. There is also a space marked “unable to respond.”

In particular, the categories “MANAGE DIVERSITY” and “BUILD PUBLIC TRUST” reflect a special relationship with the American public and the unique responsibilities of a law enforcement service. These categories attempt to tap these attributes since they are thought to be critical for success in law enforcement. Delattre’s (1994) work influenced the decision to include these items in the survey. He suggests that in order to lead effectively in the law enforcement domain, public trust issues are essential. In fact, this aspect of the FBI’s steadfastness is a significant part of the foundation upon which the agency built its reputation. As such, it is a critical part of why the organization continues to do its job of gathering information about some of the most important matters under the public domain including foreign intelligence and terrorist activities, and what allows it to enjoy a history of many significant prosecutorial successes.

## Procedures

The logistical process of distributing the material began with the instruments being mailed to the individual being evaluated with instructions to have one of their subordinates anonymously distribute them. The only personal information asked of the respondent was whether they were a peer, a direct report or the evaluated individual. Specific details were provided when the instruments were distributed, which included mailing instructions.<sup>11</sup> In addition to the ratings based on the Likert scale, survey respondents were given the opportunity to provide narrative comments.<sup>12</sup> Once completed, each rater sent their form individually to an independent data collection center that handled the data collection process.

---

<sup>10</sup> The numbers 1 – 5 are used to correspond with these responses. 1 represents the lowest measure, “highly dissatisfied,” and 5 represents the highest response, “highly satisfied” with the numbers 2, 3, and 4 representing the corresponding middle responses.

<sup>11</sup> See Appendix D.

<sup>12</sup> The narrative comments are beyond the scope of this particular research, but can be useful when providing feedback.

At the collection center the information was entered into a database and a personal report was generated for each of the supervisors on whom the information was gathered. As noted, the individual's report was considered completely confidential.

The contractor forwarded the gathered data to the researcher. It was screened carefully to ensure the data sets contained reasonable responses, and that evaluators had followed instructions. The Statistical Package for the Social Sciences (SPSS) software was used for data entry. Data from 73 individuals were used for this research, and represented all the available data through 2002. All data analysis was done throughout the fall of 2002 and spring, summer and fall of 2003.

### Research Design and Analysis

The overall research design corresponds to obtaining three types of data (self, peers and subordinates, the latter two types considered "others" when grouped together) at two points in time. At each time, the same self-administered instrument or inventory of questions (the LCCIS) dealing with leadership behavior was used. The inventory's questions were devised to fall into 11 conceptually different categories or dimensions, which are discussed along with information about their reliability in the "Instrument" section above. The overall research design broadly corresponds to a 3 x 2 x 11 design. That is, three types of individuals x two points in time x 11 sub-categories of scores gathered from each type of person.<sup>13</sup>

SPSS was selected as the system for data analysis. The original data were screened for missing values to ensure that only those cases were retained where raters took the assessment seriously and conscientiously completed the questionnaire. Extra sets of data provided with the original were screened and eliminated. Only data sets matched with the same individual at two points in time were retained. Seventy-three individuals with both a first and second assessment, therefore totaling 146 cases, were retained. If surveys were found to be missing some data points in an apparent random configuration, the data were estimated utilizing procedures outlined by Tabachnick and Fidell (1996).

---

<sup>13</sup> Additionally, adding together all the variables included in the 11 sub-categories created an overall measure.

The 146 (73 x 2) cases included a total of 1,014 subordinate, peer, and self-ratings at time one and 934 subordinate, peer, and self-ratings at time two for a total of 1,948. For each of these ratings there were 97 variables.<sup>14</sup>

Two SPSS files were originally created, each containing 73 cases. The first represented feedback at time one assessments and the second represented time two-assessments on the same individual. A six-digit sequence number was used to identify each case. Therefore, each of the 73 individuals used in the study had two six-digit numbers corresponding to two sets of data, one for each time assessment. Cases were further broken down into three categories of self-scores, peer scores and subordinate scores. A system for comparison of the scores assessed at time one (prior to EDI I) against the scores at time two (prior to EDI II) was developed. Basic descriptive statistics of mean, range, and standard deviation of group responses were analyzed.

A third SPSS file was created including scores of self, peers, subordinates, and others from each of the subscales and an overall scale at time one and time two.<sup>15</sup> Thus, this file included 73 cases broken down into the 11 conceptually different sub-categories or dimensions plus an overall measurement for self, peers, subordinates, and others at two times.<sup>16</sup> This file was used to run Analysis of Covariance (ANCOVA) tests.

Finally, a fourth SPSS file was created which was used for t tests and allowed for comparisons of significant difference of means between first and second assessments. These tests were done for each hypothesis, all sub-hypotheses, and for direct comparison of subordinate and peer mean differences at time one versus time two even though these comparisons were beyond the scope of the offered hypotheses.

Standard analytical procedures and alpha reliability analysis were conducted to evaluate the reliability of the LCCIS survey items and integrity of scales.<sup>17</sup> The results were very similar to those found by the Leadership Research Institute at the University of Connecticut, and indicated exceptionally strong category reliabilities ranging from .84 to .97. Large sample tests to compare group means (Z scores) and t tests were used as the

---

<sup>14</sup> Thus, 188,956 individual data points were originally created.

<sup>15</sup> Means were used for subordinate, peers and others.

<sup>16</sup> Thus, this file included 7008 data points (73 cases x 2 times x 12 variables x 4 types: self, peers, subordinates, and others).

<sup>17</sup> See Appendix E.

primary way of testing the research hypotheses.<sup>18</sup> Additionally, ANCOVA was used to augment the findings.

Individual hypotheses were examined as follows. Hypothesis 1 was tested using an analysis of the overall data collected by the multi-rater feedback instrument used by the FBI. Because this was a general hypothesis, all 97 questions of the LCCIS in both the self and others (subordinate and peer) groups were pooled to yield an overall scale for the entire instrument, and a comparison was made of first time assessments to second time assessments. Then a Z score was computed for comparison of the significance of difference of means at time one and time two. T tests were run for comparison of the significant difference of means between first and second assessments, and ANCOVA techniques were used to augment the findings. Additionally, while not part of the hypothesis a Z score also was computed to compare means of others' scores at time one with means of self scores at time two.

Hypotheses 2, 3 and 4 including Sub-hypotheses 2a – 2k, 3a – 3k, and 4a – 4k were all tested in the same way as the first hypothesis. As such, a significant amount of data were generated from this testing, which allowed for comparisons between leadership dimensions and gave greater depth to this study.

As with earlier analysis, similar tests were run for Hypothesis 5, even though this hypothesis was only examining a general finding rather than each of the sub-scales. Because of the broader nature of Hypothesis 5, Z score testing was not run on each of the sub-scales as was done in earlier hypotheses. However, all the other tests were run. The results are included in Chapter 4 of this study.

## Conclusion

This chapter described the methods and procedures used to obtain the data necessary to address the research questions offered. It began with the design, setting, and description of the participants and raters. It followed by including a description of the multi-rater instrument used in the present research. Finally, it concluded with the procedure and data analysis.

---

<sup>18</sup>Walker and Lev (1969) indicate the appropriateness of using difference of means to test a hypothesis that two means are equal. While difference of means yielded suitable results for the way the hypotheses were offered, ANCOVA testing was also done to account for effects of first assessments.

The next chapter will analyze the data, focusing on the phenomena of interest. Finally, Chapter 5 will provide the discussion, summary and implications of this research. Additionally, it will describe the limitations of the analysis, and conclude with recommendations for further study and practice. Between these two chapters a deeper appreciation of the concepts addressed in this research will become apparent, providing an increased understanding of the use of multi-rater feedback as a developmental tool.

## CHAPTER 4

### Research Findings

#### Introduction

This chapter discusses the findings and results developed from this research. It begins with the purpose of the study and is followed by an instrument summary and data analysis. The data analysis details the hypotheses used and is ordered by a presentation of the results of the data framed around each. Through this process a generous amount of information about multi-rater feedback was developed, the implications of which will be discussed in Chapter 5.

#### Purpose

The purpose of this study was to answer the following questions regarding multi-rater feedback in the context of the FBI work setting.

1. Is there a greater perceptual agreement between others (peers and subordinates) and self on the second multi-rater assessment than the initial assessment?
2. What are the dimensions (sub-categories of the LCCIS) of any differences found?
3. Are there differences in agreement between peer and self-ratings and subordinate and self-ratings at time two and time one?
4. What are the dimensions of any differences found in peer and self-ratings?
5. What are the dimensions of any differences found in subordinate and self-ratings?
6. Are there differences in agreement between peer and subordinate ratings at time one and time two?

These questions were answered through five hypotheses, which included a number of related sub-hypotheses. Examining them resulted in further knowledge of the multi-rater feedback process, particularly as it relates to law enforcement supervisors in the FBI. The way the Research Questions correspond to the Hypotheses is depicted in Table 1.



Table 1

### Research Questions and Corresponding Hypotheses

Research Question	Corresponding Hypotheses
Question 1	Hypothesis 1
Question 2	Hypothesis 2
Question 3	Hypotheses 3 & 4
Question 4	Hypothesis 3
Question 5	Hypothesis 4
Question 6	Hypothesis 5

### Instrument Summary

As articulated in Chapter 3, this study made use of the Leadership Commitments and Credibility Inventory System (LCCIS). Previous validation of this instrument indicated that overall values of alpha reliability were high. However, reliability tests were also conducted here to verify these previous results.<sup>19</sup> With regard to these tests, results showed that the overall value alpha reliabilities were .84 or better for each of the subscales, and were consistent with the previous validity and reliability tests reported by the Leadership Research Institute at the University of Connecticut, which are attached as Appendix C.<sup>20</sup>

### Data Analysis

To set the stage for understanding the results presented here a series of statistical tests of significance, one for each hypothesis, was used to test whether various sets of ratings were relatively closer to each other at time two, compared with time one. Each statistical test was evaluated for statistical significance in the same way, as follows:

- Z scores were computed, to compare average values of each group at the same point in time.
- Each Z score was evaluated to see if its value was greater than or equal to 1.96. This represents a criterion value for considering differences in group means as not likely to have occurred by chance less than five times in 100 (corresponding to an alpha level of .05). However, if a series of related sub-hypotheses (as in Research Questions 2, 3, 4, and 5 where 11 categories from the same instrument were being examined) was

---

<sup>19</sup> See Appendix E

<sup>20</sup> Reliability tests were conducted only on time one assessments.

tested by using data from the same persons, then the significance tests were considered interrelated, rather than having independent or separate probability estimates of how likely they were to be significant. The overall criterion value of .05 for these interrelated tests was divided by 11, the number of sub-hypotheses, to make the alpha level .0045 for each of the 11 tests. This corresponds to a Z score of 2.84. [This adjustment (developed by Dunn) ensures that the overall probability level for the total number of related significance tests does not exceed .05. In that way, it reduces the chance of “too many” tests being considered statistically significant, when in fact they are not because of their being tested with data from the same sources.]<sup>21</sup>

- Each statistical test was considered two-tailed, since each hypothesis tested for whether or not one group’s results was larger or smaller than results from the other group, rather than testing for the direction of those differences.
- To understand the results better, disparities between group means were interpreted after conducting statistical tests, as follows. If a statistical test was found to be statistically significant (that is, its Z score exceeding or being equal to the critical value and therefore falling in the critical region), then the group means involved were subtracted from each other, to obtain a numeric difference. This was done to see how numerically far apart they were in contributing to the statistical disparity, and in what direction. If a test result was not statistically significant, then the group mean differences compared in that test were considered numerically different but not significantly so. Instead, they were considered chance rather than meaningful differences.
- Finally, t tests were run for comparison of the significant difference of means between first and second assessments.<sup>22</sup>

After tests for significance were done, ANCOVA was used to further test each hypothesis. Since there was variation in initial values, time one scores were used as the

---

<sup>21</sup> See Dunn in Kirk, R.E. (1968). *Experimental Design: Procedures for the Behavioral Sciences*. Belmont, CA: Brooks/Cole Publishing.

<sup>22</sup> It is noted that in every hypotheses and sub-hypotheses t tests examining difference of means between first and second assessments yielded non-significant alpha levels. As such, the following results are examined with the understanding that they are relative ones regarding the raw data. They add depth to this research which would be lost if this study focused only on significant alpha levels.

covariate in order to determine whether response type (peer, subordinate, or others) makes a difference in the final response (time two) when time one scores were controlled for. All ANCOVA test were computed using alpha = .05. These tests added robustness to the findings of the tests of significance that determined the level of agreement between self-ratings and others' ratings.

### *Hypotheses*

Most multi-source research on the level of agreement between self-ratings and others' ratings has found generally little agreement between the two (Atwater & Yammarino, 1992). Further, there is evidence that individuals tend to overrate themselves as compared to others (Thornton, 1980; Harris & Schaubroeck, 1988). The present research examined the following hypotheses related to these research findings, but within the context of the FBI work setting.

### *Hypothesis 1*

The present research builds on the former findings. Here it was hypothesized that having two sets of ratings between self and others at different times would bring about more alignment on the second set. Therefore, the first hypothesis considered the agreement between self-perception and perception of others on second assessments as compared to first assessments, and sought to answer the initial question of the research, which is:

1. Is there a greater perceptual agreement between others (peers and subordinates) and self on the second multi-rater assessment than the initial assessment?

**Hypothesis 1: Overall, leaders in the FBI will rate their leadership more highly compared with others (peers and subordinates) on first assessments than on second assessments [computed as Self Mean Assessments (SA) – Other Mean Assessments (OA) at time one (T1) > Self Mean Assessments – Other Mean Assessments at time two (T2)].**

To test this hypothesis, all measures were combined to create an overall or global leadership scale. This was used to compare first assessments' means of self to others (peers and subordinates) and second assessments' means of self to others. This process included a decision to equalize the importance of the values and not have a weighted average from one or more of the subscales that make up the different categories of the

LCCIS. As such, each of the 97 questions that compose the instrument was given the same weight.

First, an overall mean of all 73 cases of self-scores at time one was computed and compared to an overall mean score of all others (peers and subordinates) at time one. An overall mean of all 73 cases of self-scores at time two also was computed and compared to an overall mean score of all others at time two (see Table 2). Z scores were then computed to compare means at time one and those at time two, in order to test the null hypothesis that there is no difference in means in either instance. The observed value of  $Z = 1.34$  at time one, and the observed value of  $Z = .09$  at time two were not within the rejection region. Therefore, there was insufficient evidence to reject the null hypothesis in each case. Consequently, the results suggest no statistical difference between perceptual agreement measures of self versus others at time one, or at time two. This was confirmed by a t test comparing difference of means at time one with time two.

While not part of the hypothesis a Z score also was computed to compare means of others' scores at time one with means of self scores at time two, to examine if there was a statistically significant difference. The observed value of  $-1.66$  was also not within the rejection region. This indicates that no significant differences were found in perceptual agreement measures between others and self at time one versus time two. This analysis adds further depth to the overall finding that no significant differences were found in perceptual agreement measures at time one versus time two for this hypothesis.

It is noted, however, that while these Z score computations were not found to be significant, the patterns in mean differences are of interest. Although not significant, the numerical difference in means ( $.0203$  at time one  $> .0016$  at time two) suggest directional consistency relative to Hypothesis 1 in smaller differences being associated with greater feedback. These implications will be discussed further in Chapter 5.

Table 2

Descriptive Statistics: Hypothesis 1 (Overall Measurement)

	N	Mean	Std.Deviation
OVERALL Self Time One	73	4.4324	.30176
OVERALL Others Time One	941	4.4121	.57332
OVERALL Self Time Two	73	4.4403	.33565
OVERALL Others Time Two	861	4.4387	.58740

To further test this hypothesis ANCOVA was used. As with the Z score testing, all measures were combined to create an overall or global leadership scale by adding together all the data points in the 11 sub-categories and creating an overall mean of others (peers and subordinates). This was done for each of the 73 cases at time one and time two. The between-subjects factors of 73 mean self scores were then compared with 73 mean other scores. The dependent variable was the mean overall scale at time two and the covariate was the mean overall scale at time one. Results showed a significance level of .94 indicating a non-significant alpha level. It adds further depth to the overall finding that no significant differences were found in perceptual agreement measures at time one versus time two for this hypothesis. Table 3 reflects the between-subjects effects of this test.

Table 3

ANCOVA: Hypothesis 1 (Overall Measurement)

Tests of Between-Subjects Effects

Dependent Variable: ALL2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.956 <sup>a</sup>	2	2.478	32.828	.000
Intercept	2.314	1	2.314	30.651	.000
ALL1	4.943	1	4.943	65.479	.000
RESPTYPE	4.997E-04	1	4.997E-04	.007	.935
Error	10.794	143	7.548E-02		
Total	2877.620	146			
Corrected Total	15.750	145			

a. R Squared = .315 (Adjusted R Squared = .305)

*Hypothesis 2*

While the first hypothesis examined an overall measure, it may not convey the entire story. In order to more carefully examine the subcomponents of the LCCIS, the second hypothesis deals with the nature of the change in agreement between self-ratings and others' ratings on second assessments as compared to first assessments as reflected in various types of dimensions or subcategories of the leadership assessment. It seeks to answer the following research question:

2. What are the dimensions (sub-categories of the LCCIS) of any differences found?

**Hypothesis 2: Leaders in the FBI will rate their leadership more highly compared with others (peers and subordinates) on first assessments than on second assessments in each subcategory [computed as Self Mean Assessments (SA) – Other Mean Assessments (OA) at time one (T1) > Self Mean Assessments – Other Mean Assessments at time two (T2) in each subcategory].**

In order to test this hypothesis, 11 Sub-Hypotheses 2a – 2k were offered. These sub-hypotheses allowed for an examination of each subscale within the instrument that composes the LCCIS. A lack of support for any of the sub-hypotheses surrounding these subscales indicated a lack of overall support for Hypothesis 2. These sub-hypotheses were all tested in the same way as Hypothesis 1. All the sub-hypotheses are listed below, but only those with noteworthy findings are discussed in any detail.

**Hypothesis 2a: There will be increased agreement between self and others on second assessments in the category “STEPS to EMPOWERMENT” compared with first assessments.**

There was no support for this hypothesis. Table 4 reflects the mean scores of this subcategory at time one and time two.

Table 4

Descriptive Statistics: Hypothesis 2a (Steps to Empowerment)

	N	Mean	Std. Deviation
STOE Self Time One	73	4.2952	.32619
STOE Others Time One	941	4.3016	.65725
STOE Self Time Two	73	4.3438	.33218
STOE Others Time Two	861	4.3539	.64907

As with Hypothesis 1, ANCOVA testing was also done. Results showed a non-significant alpha level of .98. It adds further depth to the overall finding that no significant differences were found in perceptual agreement measures at time one versus time two for this hypothesis. Table 5 reflects the between-subjects effects of this test.

Table 5

ANCOVA: Hypothesis 2a (Steps to Empowerment)

**Tests of Between-Subjects Effects**

Dependent Variable: STOE2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.143 <sup>a</sup>	2	2.071	23.857	.000
Intercept	4.776	1	4.776	55.008	.000
STOE1	4.143	1	4.143	47.714	.000
RESPTYPE	6.505E-05	1	6.505E-05	.001	.978
Error	12.416	143	8.683E-02		
Total	2771.420	146			
Corrected Total	16.559	145			

a. R Squared = .250 (Adjusted R Squared = .240)

**Hypothesis 2b: There will be increased agreement between self and others on second assessments in the category “CREDIBILITY” compared with first assessments.**

The observed value of  $Z = 3.66$  at time one is within the rejection region. However, the observed value of  $Z = 1.31$  at time two is not. The results at time two showed no statistical difference between the means of self versus others. Therefore, there is increased agreement between self and others on second assessments in the category CREDIBILITY as compared to first assessments and support for Hypothesis 2b. Table 6 reflects the mean scores of this subcategory at time one and at time two.



Table 6

Descriptive Statistics: Hypothesis 2b (Credibility)

	N	Mean	Std. Deviation
CRED Self Time One	73	4.4514	.33980
CRED Others Time One	941	4.3774	.67953
CRED Self Time Two	73	4.4292	.35710
CRED Others Time Two	861	4.4003	.68675

ANCOVA testing was done to further test this hypothesis. Results showed a non-significant alpha level of .99. Table 7 reflects the between-subjects effects of this test.

Table 7

ANCOVA: Hypothesis 2b (Credibility)

Tests of Between-Subjects Effects

Dependent Variable: CRED2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.865 <sup>a</sup>	2	3.432	37.344	.000
Intercept	2.883	1	2.883	31.372	.000
CRED1	6.771	1	6.771	73.668	.000
RESPTYPE	1.044E-05	1	1.044E-05	.000	.992
Error	13.144	143	9.191E-02		
Total	2850.970	146			
Corrected Total	20.008	145			

a. R Squared = .343 (Adjusted R Squared = .334)

**Hypothesis 2c: There will be increased agreement between self and others on second assessments in the category “CHALLENGE and SUPPORT the ORGANIZATION” compared with first assessments.**

Z scores for this hypothesis are – 4.07 and – 2.75 at time one and time two, respectively. The time one score is within the rejection region whereas the time two

score is not. Similar to Hypothesis 2b, there is increased agreement between self and others on second assessments in the category CHALLENGE and SUPPORT the ORGANIZATION as compared to first assessments and support for Hypothesis 2c.

It is curious to note that for this subcategory the supervisors underrated themselves as compared to others at time one and time two. This is reflected in the negative differences in mean scores,  $-.074$  at time one and  $-.069$  at time two. The decrease in difference of mean scores at time two shows a convergence of self and others' assessments from time one and, therefore, increased agreement, although it is not in the direction of supervisors overrating themselves at time one as was presumed when this study was proposed. See Table 8 for mean scores for this hypothesis.

Table 8

Descriptive Statistics: Hypothesis 2c (Challenge and Support the Organization)

	N	Mean	Std. Deviation
CSORG Self Time One	73	4.5137	.37035
CSORG Others Time One	941	4.5886	.52638
CSORG Self Time Two	73	4.5114	.44125
CSORG Others Time Two	861	4.5811	.56933

ANCOVA testing was also done. Results showed a non-significant alpha level of .72. Table 9 reflects the between-subjects effects of this test.

Table 9

ANCOVA: Hypothesis 2c (Challenge and Support the Organization)

**Tests of Between-Subjects Effects**

Dependent Variable: CSO2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.681 <sup>a</sup>	2	2.341	21.661	.000
Intercept	3.124	1	3.124	28.911	.000
CSO1	4.549	1	4.549	42.094	.000
RESPTYPE	1.430E-02	1	1.430E-02	.132	.717
Error	15.452	143	.108		
Total	3030.880	146			
Corrected Total	20.133	145			

a. R Squared = .233 (Adjusted R Squared = .222)

**Hypothesis 2d: There will be increased agreement between self and others on second assessments in the category “BUILD TEAMS” compared with first assessments.**

There was no support for this hypothesis. Table 10 reflects the mean scores of this subcategory at time one and time two.

Table 10

**Descriptive Statistics: Hypothesis 2d (Build Teams)**

	N	Mean	Std. Deviation
BT Self Time One	73	4.3562	.48218
BT Others Time One	941	4.3964	.69249
BT Self Time Two	73	4.4247	.52156
BT Others Time Two	861	4.4051	.71614

ANCOVA results showed a non-significant alpha level of .44. It adds further depth to the overall finding that no significant differences were found for this hypothesis. Table 11 reflects the between-subject effects of this test.

Table 11  
ANCOVA: Hypothesis 2d (Build Teams)

**Tests of Between-Subjects Effects**

Dependent Variable: BT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.468 <sup>a</sup>	2	2.734	15.891	.000
Intercept	8.757	1	8.757	50.899	.000
BT1	5.418	1	5.418	31.492	.000
RESPTYPE	.101	1	.101	.590	.444
Error	24.602	143	.172		
Total	2871.610	146			
Corrected Total	30.070	145			

a. R Squared = .182 (Adjusted R Squared = .170)

**Hypothesis 2e: There will be increased agreement between self and others on second assessments in the category “DEVELOP INDIVIDUALS” compared with first assessments.**

Z scores for this hypothesis are – 3.93 and .11 at time one and time two, respectively. The Z score at time one is within the rejection region however, the Z score at time two is not. Therefore, there was support for Hypothesis 2e and, like Hypothesis 2c it was not because the supervisors overrated themselves at time one as was presumed when this study was proposed. Table 12 reflects the mean scores of this subcategory at time one and time two.

Table 12

Descriptive Statistics: Hypothesis 2e (Develop Individuals)

	N	Mean	Std. Deviation
DI Self Time One	73	4.2661	.44308
DI Others Time One	941	4.3691	.62212
DI Self Time Two	73	4.4012	.41626
DI Others Time Two	861	4.3985	.64950

ANCOVA testing was done to further test this hypothesis. Results showed a non-significant alpha level of .37. Table 13 reflects the between-subjects effects of this test.

Table 13

ANCOVA: Hypothesis 2e (Develop Individuals)

Tests of Between-Subjects Effects

Dependent Variable: DI2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.141 <sup>a</sup>	2	1.071	8.243	.000
Intercept	10.515	1	10.515	80.952	.000
DI1	2.119	1	2.119	16.315	.000
RESPTYPE	.103	1	.103	.794	.374
Error	18.574	143	.130		
Total	2840.240	146			
Corrected Total	20.716	145			

a. R Squared = .103 (Adjusted R Squared = .091)

**Hypothesis 2f: There will be increased agreement between self and others on second assessments in the category “QUALITY and TASK” compared with first assessments.**

There was no support for this hypothesis. See Table 14 for mean scores for this hypothesis.

Table 14

Descriptive Statistics: Hypothesis 2f (Quality and Task)

	N	Mean	Std. Deviation
QT Self Time One	73	4.4315	.39224
QT Others Time One	941	4.4497	.61714
QT Self Time Two	73	4.4425	.43745
QT Others Time Two	861	4.4705	.63230

ANCOVA results showed a non-significant alpha level of .82. It adds further depth to the overall finding that no significant differences were found in perceptual agreement measures for this hypothesis. Table 15 reflects the between-subjects effects of this test.

Table 15

ANCOVA: Hypothesis 2f (Quality and Task)

Tests of Between-Subjects Effects

Dependent Variable: QT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3.669 <sup>a</sup>	2	1.835	13.929	.000
Intercept	6.329	1	6.329	48.049	.000
QT1	3.661	1	3.661	27.796	.000
RESPTYPE	7.034E-03	1	7.034E-03	.053	.818
Error	18.836	143	.132		
Total	2913.670	146			
Corrected Total	22.505	145			

a. R Squared = .163 (Adjusted R Squared = .151)

**Hypothesis 2g: There will be increased agreement between self and others on second assessments in the category “CUSTOMER/CLIENT” compared with first assessments.**

There was no support for this hypothesis. Table 16 reflects the mean scores of this subcategory at time one and time two.

Table 16

Descriptive Statistics: Hypothesis 2g (Customer/Client)

	N	Mean	Std. Deviation
CC Self Time One	73	4.4132	.43842
CC Others Time One	941	4.3822	.62411
CC Self Time Two	73	4.3425	.45550
CC Others Time Two	861	4.4344	.60268

ANCOVA results showed a non-significant alpha level of .13. Table 17 reflects the between-subjects effects of this test.

Table 17

ANCOVA: Hypothesis 2g (Customer/Client)

**Tests of Between-Subjects Effects**

Dependent Variable: CC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3.397 <sup>a</sup>	2	1.698	13.036	.000
Intercept	8.169	1	8.169	62.707	.000
CC1	3.197	1	3.197	24.539	.000
RESPTYPE	.300	1	.300	2.300	.132
Error	18.630	143	.130		
Total	2824.000	146			
Corrected Total	22.026	145			

a. R Squared = .154 (Adjusted R Squared = .142)

**Hypothesis 2h: There will be increased agreement between self and others on second assessments in the category “VISION” compared with first assessments.**

There was a lack of support for this hypothesis. See Table 18 for mean scores for this hypothesis.

Table 18

Descriptive Statistics: Hypothesis 2h (Vision)

	N	Mean	Std. Deviation
V Self Time One	73	4.2849	.46776
V Others Time One	941	4.3260	.72026
V Self Time Two	73	4.3096	.52709
V Others Time Two	861	4.3566	.71893



As with previous hypotheses, ANCOVA testing was also done. Results showed a non-significant alpha level of .79. It adds further depth to the overall finding that no significant differences were found for this hypothesis. Table 19 reflects the between-subjects effects of this test.

Table 19

ANCOVA: Hypothesis 2h (Vision)

**Tests of Between-Subjects Effects**

Dependent Variable: V2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.136 <sup>a</sup>	2	3.568	23.719	.000
Intercept	5.807	1	5.807	38.603	.000
V1	7.106	1	7.106	47.238	.000
RESPTYPE	1.080E-02	1	1.080E-02	.072	.789
Error	21.510	143	.150		
Total	2758.370	146			
Corrected Total	28.646	145			

a. R Squared = .249 (Adjusted R Squared = .239)

**Hypothesis 2i: There will be increased agreement between self and others on second assessments in the category “FOSTER CREATIVITY” compared with first assessments.**

There was no support for this hypothesis. Table 20 reflects the mean scores of this subcategory at time one and time two.

Table 20

Descriptive Statistics: Hypothesis 2i (Foster Creativity)

	N	Mean	Std. Deviation
FC Self Time One	73	4.4178	.51884
FC Others Time One	941	4.3448	.72528
FC Self Time Two	73	4.4315	.44733
FC Others Time Two	861	4.3673	.72037

ANCOVA testing was done to further test this hypothesis. Results showed a non-significant alpha level of .56. It adds further depth to the overall finding that no significant differences were found for this hypothesis. Table 21 reflects the between-subjects effects of this test.

Table 21

ANCOVA: Hypothesis 2i (Foster Creativity)

Tests of Between-Subjects Effects

Dependent Variable: FC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.014 <sup>a</sup>	2	3.007	20.204	.000
Intercept	9.900	1	9.900	66.524	.000
FC1	5.814	1	5.814	39.066	.000
RESPTYPE	5.149E-02	1	5.149E-02	.346	.557
Error	21.282	143	.149		
Total	2846.820	146			
Corrected Total	27.296	145			

a. R Squared = .220 (Adjusted R Squared = .209)

**Hypothesis 2j: There will be increased agreement between self and others on second assessments in the category “MANAGE DIVERSITY” compared with first assessments.**

Z scores for this hypothesis are 8.42 and 5.31 at time one and time two, respectively. Therefore, there was sufficient evidence to reject the null hypothesis in both of these cases. The mean score difference at time one (.20) is greater than the mean score difference at time two (.14). Therefore, there is a convergence of means at time two as compared to time one. Thus, there is support for Hypothesis 2j. Table 22 reflects the mean scores of this subcategory.

Table 22

Descriptive Statistics: Hypothesis 2j (Manage Diversity)

	N	Mean	Std. Deviation
MD Self Time One	73	4.7260	.41029
MD Others Time One	941	4.5248	.64481
MD Self Time Two	73	4.6849	.42416
MD Others Time Two	861	4.5461	.67381

As with previous hypotheses, ANCOVA testing was done. Results showed a non-significance alpha level of .50. Table 23 reflects the between-subjects effects of this test.

Table 23

ANCOVA: Hypothesis 2j (Manage Diversity)

**Tests of Between-Subjects Effects**

Dependent Variable: MD2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.511 <sup>a</sup>	2	3.255	28.379	.000
Intercept	4.648	1	4.648	40.520	.000
MD1	5.697	1	5.697	49.664	.000
RESPTYPE	5.322E-02	1	5.322E-02	.464	.497
Error	16.404	143	.115		
Total	3126.090	146			
Corrected Total	22.915	145			

a. R Squared = .284 (Adjusted R Squared = .274)

**Hypothesis 2k: There will be increased agreement between self and others on second assessments in the category “BUILD PUBLIC TRUST” compared with first assessments.**

Z scores for this hypothesis are 6.39 and .98 at time one and time two, respectively. The Z score at time one is within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in this case, but not at time two. A convergence of mean scores at time two also occurred as compared with time one. Thus, there is support for Hypothesis 2k. Table 24 reflects the mean scores of this subcategory at time one and time two.

Table 24

Descriptive Statistics: Hypothesis 2k (Build Public Trust)

	N	Mean	Std. Deviation
BPT Self Time One	73	4.7412	.32450
BPT Others Time Two	941	4.6500	.47157
BPT Self Time Two	73	4.6743	.40964
BPT Others Time Two	861	4.6531	.51165

ANCOVA testing was done to further test this hypothesis. Results showed a non-significant alpha level of .82. Table 25 reflects the between-subjects effects of this test.

Table 25

ANCOVA: Hypothesis 2k (Build Public Trust)

Tests of Between-Subjects Effects

Dependent Variable: BPT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.459 <sup>a</sup>	2	1.229	11.644	.000
Intercept	3.298	1	3.298	31.234	.000
PBT1	2.416	1	2.416	22.882	.000
RESPTYPE	5.725E-03	1	5.725E-03	.054	.816
Error	15.099	143	.106		
Total	3183.750	146			
Corrected Total	17.558	145			

a. R Squared = .140 (Adjusted R Squared = .128)

Because there was a lack of support for six sub-hypotheses (2a, 2d, 2f, 2g, 2h, and 2i) and t tests showed a non-significant results in comparing time one difference of means to time two difference of means in all sub-categories, there was a lack of support for Hypothesis 2. However, by testing these sub-hypotheses, valuable information regarding the 11 subscales measured was developed. The implications will be discussed in Chapter 5.

### *Hypothesis 3*

Consistent with the second hypothesis, Hypothesis 3 flows from the primary hypothesis. This hypothesis, together with Hypothesis 4, separates “others” into peers and subordinates. These two hypotheses therefore assist in addressing the third, fourth, and fifth research questions:

3. Are there differences in agreement between peer and self-ratings and subordinate and self-ratings at time two and time one?
4. What are the dimensions of any differences found in peer and self-ratings?
5. What are the dimensions of any differences found in subordinate and self-ratings?

Specifically, Hypothesis 3, addresses part of the third research question, the fourth research question fully, and separates “others” (from Hypothesis 1) into the portion made up of peers. It is as follows:

**Hypothesis 3: Leaders in the FBI will rate their leadership more highly compared with peers (PA) on first assessments than on second assessments in each subcategory [SA – PA at T1 > SA – PA at T2 in each subcategory].**

As with Hypothesis 2, in order to test this hypothesis, Sub-Hypotheses 3a – 3k were offered. These sub-hypotheses allowed each subscale within the LCCIS to be examined. A lack of support for any of the sub-hypotheses surrounding these subscales indicated a lack of overall support for Hypothesis 3. These sub-hypotheses were all tested in the same way as Hypotheses 1 and 2. All the sub-hypotheses are listed below, but only those with noteworthy findings are discussed in any detail.

As with Hypothesis 2, ANCOVA test on a global or overall leadership scale is included comparing between-subjects factors of 73 mean self scores with 73 mean peer scores after Hypothesis 3k. Additionally, ANCOVA testing was done in the same way on each of these subscales to add greater understanding to this research. In every ANCOVA except the CUSTOMER/CLIENT sub-category results showed a non-significant alpha level. Therefore, only with Hypothesis 3g, which deals with the CUSTOMER/CLIENT sub-category, are the results of ANCOVA included in the data analysis of Sub-Hypotheses 3a – 3k. <sup>23</sup>

---

<sup>23</sup> See Appendix F for the entire results of Hypothesis 3 ANCOVA testing.

**Hypothesis 3a: There will be increased agreement between self and peers on second assessments in the category “STEPS to EMPOWERMENT” compared with first assessments.**

Z scores for this hypothesis are – 6.51 and – 4.74 at time one and time two, respectively. Both are within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. This hypothesis is similar to Sub-Hypothesis 2c where the supervisors underrated themselves as compared with others. In this instance, as in most of the subcategories for Hypothesis 3, the supervisors underrated themselves as compared to their peers at time one and time two. Presently, this is reflected in the negative difference in mean scores, – .12 at time one and – .10 at time two. The decrease in difference of mean scores at time two shows a convergence of self and peers’ assessments from time one versus time two and therefore increased agreement. Therefore, there is support for this hypothesis, although it is not because the supervisors overrated themselves at time one as was presumed when this study was proposed. Table 26 reflects the mean scores of this subcategory at time one and time two.

Table 26

Descriptive Statistics: Hypothesis 3a (Steps to Empowerment)

	N	Mean	Std. Deviation
STOE Self Time One	73	4.2952	.32619
STOE Peers Time One	442	4.4149	.53355
STOE Self Time One	73	4.3438	.33218
STOE Peers Time Two	386	4.4456	.58166

**Hypothesis 3b: There will be increased agreement between self and peers on second assessments in the category “CREDIBILITY” compared with first assessments.**

There was a lack of support for this hypothesis. See Table 27 for mean scores for this hypothesis.

Table 27

Descriptive Statistics: Hypothesis 3b (Credibility)

	N	Mean	Std. Deviation
CRED Self Time One	73	4.4514	.33980
CRED Peers Time One	442	4.5157	.51211
CRED Self Time Two	73	4.4292	.35710
CRED Peers Time Two	386	4.4996	.61179

**Hypothesis 3c: There will be increased agreement between self and peers on second assessments in the category “CHALLENGE and SUPPORT the ORGANIZATION” compared with first assessments.**

Z scores for this hypothesis are – 8.85 and – 4.80 at time one and time two, respectively. Both are within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. This hypothesis is similar to Sub-Hypothesis 3a where the supervisors underrated themselves as compared to peers at time one and time two. Presently, this is reflected in the negative difference in mean scores: – .16 at time one and – .13 at time two. The decrease in difference of mean scores at time two shows a convergence of self and peers’ assessments from time one to time two and, therefore, increased agreement. See Table 28 for mean scores for this hypothesis.



Table 28

Descriptive Statistics: Hypothesis 3c (Challenge and Support the Organization)

	N	Mean	Std. Deviation
CSORG Self Time One	73	4.5137	.37035
CSORG Peers Time One	442	4.6765	.43633
CSORG Self Time Two	73	4.5114	.44125
CSORG Peers Time Two	386	4.6429	.54463

**Hypothesis 3d: There will be increased agreement between self and peers on second assessments in the category “BUILD TEAMS” compared with first assessments.**

Z scores for this hypothesis are – 5.47 and – 1.98 at time one and time two, respectively. Therefore, there was insufficient evidence to reject the null hypothesis at time one but not at time two. This hypothesis is similar to Sub-Hypotheses 3a and 3c where the supervisors underrated themselves as compared to peers at time one and time two. Presently, this is reflected in the negative difference in mean scores: – .17 at time one and – .08 at time two. The decrease in difference of mean scores at time two shows a convergence of self and peers’ assessments from time one to time two and, therefore, increased agreement. Thus, there is support for this hypothesis. See Table 29 for mean scores for this hypothesis.

Table 29

Descriptive Statistics: Hypothesis 3d (Build Teams)

	N	Mean	Std. Deviation
BT Self Time One	73	4.3562	.48218
BT Peers Time One	442	4.5271	.56954
BT Self Time Two	73	4.4247	.52156
BT Peers Time Two	386	4.5032	.68269

**Hypothesis 3e: There will be increased agreement between self and peers on second assessments in the category “DEVELOP INDIVIDUALS” compared with first assessments.**

Z scores for this hypothesis are – 7.4 and – 3.43 at time one and time two, respectively. Both are within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. This hypothesis is similar to Sub-Hypotheses 3a, 3c, and 3d where the supervisors underrated themselves as compared to peers at time one and time two. In this instance it is reflected in the negative difference in mean scores: – .19 at time one and – .09 at time two. The decrease in difference of mean scores at time two shows a convergence of self and peers’ assessments and, therefore, increased agreement. Thus, there is support for this hypothesis. See Table 30 for mean scores.

Table 30

Descriptive Statistics: Hypothesis 3e (Develop Individuals)

	N	Mean	Std. Deviation
DI Self Time One	73	4.2661	.44308
DI Peers Time One	442	4.4586	.50116
DI Self Time Two	73	4.4012	.41626
DI Peers Time Two	386	4.4907	.56933

**Hypothesis 3f: There will be increased agreement between self and peers on second assessments in the category “QUALITY and TASK” compared with first assessments.**

Z scores for this hypothesis are  $-5.53$  and  $-3.83$  at time one and time two, respectively. Both are within the rejection region. This hypothesis is similar to Sub-Hypotheses 3a, 3c, 3d, and 3e where the supervisors underrated themselves as compared to peers at time one and time two. Presently, this is reflected in the negative difference in mean scores:  $-.12$  at time one and  $-.11$  at time two. The decrease in difference of mean scores at time two shows a convergence of self and peers’ assessments from time one to time two and, therefore, increased agreement and support for this hypothesis. See Table 31 for mean scores for this hypothesis.

Table 31

Descriptive Statistics: Hypothesis 3f (Quality and Task)

	N	Mean	Std. Deviation
QT Self Time One	73	4.4315	.39224
QT Peers Time One	442	4.5529	.51389
QT Self Time Two	73	4.4425	.43745
QT Peers Time Two	386	4.5521	.59308

**Hypothesis 3g: There will be increased agreement between self and peers on second assessments in the category “CUSTOMER/CLIENT” compared with first assessments.**

There was a lack of support for this hypothesis. See Table 32 for mean scores for this hypothesis.

Table 32

Descriptive Statistics: Hypothesis 3g (Customer/Client)

	N	Mean	Std. Deviation
CC Self Time One	73	4.4132	.43842
CC Peers Time One	442	4.4672	.55984
CC Self Time Two	73	4.3425	.45550
CC Peers Time Two	386	4.4922	.60928

As mentioned earlier, ANCOVA testing for this sub-hypothesis showed a significant alpha level (.05). This was the only sub-category dealing with ANCOVA

testing in Hypotheses 3a – 3k where a significant alpha level was found. Table 33 reflects the between-subjects effects of this test.

Table 33  
ANCOVA: Hypothesis 3g (Customer/Client)

**Tests of Between-Subjects Effects**

Dependent Variable: CC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.650 <sup>a</sup>	2	2.825	18.765	.000
Intercept	8.323	1	8.323	55.283	.000
CC1	4.924	1	4.924	32.704	.000
RESPTYPE	.604	1	.604	4.009	.047
Error	21.529	143	.151		
Total	2872.250	146			
Corrected Total	27.180	145			

a. R Squared = .208 (Adjusted R Squared = .197)

**Hypothesis 3h: There will be increased agreement between self and peers on second assessments in the category “VISION” compared with first assessments.**

Z scores for this hypothesis are – 4.84 and – 3.46 at time one and time two, respectively. Both are within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. This hypothesis is similar to the five Sub-Hypotheses 3a, 3c, 3d, 3e, and 3f where the supervisors underrated themselves as compared to peers at time one and time two. Presently, this is reflected in the negative difference in mean scores: – .16 at time one and – .14 at time two. The decrease in difference of mean scores at time two shows a convergence of self and peers’ assessments from time one to time two and, therefore, increased agreement. Therefore, there is support for this hypothesis. See Table 34 for mean scores for this hypothesis.

Table 34

Descriptive Statistics: Hypothesis 3h (Vision)

	N	Mean	Std. Deviation
V Self Time One	73	4.2849	.46776
V Peers Time One	442	4.4403	.63782
V Self Time Two	73	4.3096	.52709
V Peers Time Two	386	4.4477	.67336

**Hypothesis 3i: There will be increased agreement between self and peers on second assessments in the category “FOSTER CREATIVITY” compared with first assessments.**

Results indicated a lack of support for this hypothesis. See Table 35 for mean scores for this hypothesis.

Table 35

Descriptive Statistics: Hypothesis 3i (Foster Creativity)

	N	Mean	Std. Deviation
FC Self Time One	73	4.4178	.51884
FC Peers Time One	442	4.4106	.63638
FC Self Time Two	73	4.4315	.44733
FC Peers Time Two	386	4.4236	.68518

**Hypothesis 3j: There will be increased agreement between self and peers on second assessments in the category “MANAGE DIVERSITY” compared with first assessments.**

Z scores for this hypothesis are 2.97 and .88 at time one and time two, respectively. The Z score at time one is within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in this case, but not at time two. Further, there is a convergence of mean scores at time two as compared to time one. Thus, there is support for Hypothesis 3j. See Table 36 for mean scores for this hypothesis.

Table 36

Descriptive Statistics: Hypothesis 3j (Manage Diversity)

	N	Mean	Std. Deviation
MD Self Time One	73	4.7260	.41029
MD Peers Time One	442	4.6584	.49205
MD Self Time Two	73	4.6849	.42416
MD Peers Time Two	386	4.6617	.55720

**Hypothesis 3k: There will be increased agreement between self and peers on second assessments in the category “BUILD PUBLIC TRUST” compared with first assessments.**

There was a lack of support for this hypothesis. See Table 37 for mean scores for this hypothesis.

Table 37

Descriptive Statistics: Hypothesis 3k (Build Public Trust)

	N	Mean	Std. Deviation
BPT Self Time One	73	4.7412	.32450
BPT Peers Time One	442	4.7589	.36451
BPT Self Time Two	73	4.6743	.40964
BPT Peers Time Two	386	4.7522	.45670

Finally, to further examine Hypothesis 3, ANCOVA was used on a global leadership scale by adding together all data points in the 11 sub-categories and creating an overall mean of peers. This was done for each of the 73 cases at time one and time two. The between-subjects factors of 73 mean self scores were then compared with 73 mean peer scores. The dependent variable was the mean overall scale at time two and the covariate was the mean overall scale at time one. Results showed a non-significant alpha level of .31. It suggests a lack of support for this hypothesis from a global perspective, which is also confirmed by the findings of four of the sub-hypotheses and the t tests. Table 38 reflects the between-subjects effects of this test.



Table 38

## ANCOVA: Hypothesis 3 (Overall Measurement)

## Tests of Between-Subjects Effects

Dependent Variable: ALL2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.057 <sup>a</sup>	2	3.029	37.650	.000
Intercept	2.128	1	2.128	26.456	.000
ALL1	5.712	1	5.712	71.008	.000
RESPTYPE	8.210E-02	1	8.210E-02	1.021	.314
Error	11.503	143	8.044E-02		
Total	2955.190	146			
Corrected Total	17.560	145			

a. R Squared = .345 (Adjusted R Squared = .336)

Because there was a lack of support for four sub-hypotheses (3b, 3g, 3i, and 3k), and t tests showed a non-significant results in comparing time one difference of means to time two difference of means in all sub-categories, there was also a lack of support for Hypothesis 3. However, as in the case of Hypothesis 2, by testing these sub-hypotheses, valuable information regarding the subscales measuring 11 subcategories of the LCCIS was developed. The implications will be discussed in Chapter 5.

#### *Hypothesis 4*

Consistent with the second and third hypotheses, Hypothesis 4 also flows from Hypothesis 1. This hypothesis separates “others” (from the first hypothesis) into the portion that is made up of subordinates. Like Hypothesis 3, it assists in addressing a part of the third research question. Additionally, it addresses the fifth research question.

These questions are as follows:

3. Are there differences in agreement between peer and self-ratings and subordinate and self-ratings at time one and time two?
5. What are the dimensions of any differences found in subordinate and self ratings?

**Hypothesis 4: Leaders in the FBI will rate their leadership more highly compared with subordinates (sa) on first assessments than on second assessments in each subcategory [SA – sa at T1 > SA – sa at T2 in each subcategory].**

In order to test this hypothesis, Hypotheses 4a – 4k were offered. These hypotheses allowed each of the LCCIS subscales to be examined. A lack of support for any of the sub-hypotheses surrounding these subscales indicated a lack of overall support for Hypothesis 4. These sub-hypotheses were all tested in the same way as Hypotheses 1, 2, and 3. All the sub-hypotheses are listed below, but only those with noteworthy findings are discussed in any detail.

As with Hypotheses 2 and 3, an ANCOVA test on a global or overall leadership scale was included after Hypothesis 4k. Additionally, ANCOVA testing was done in the same way on each of these subscales to add greater understanding to this research. In every case for this hypothesis the ANCOVA showed a non-significant alpha level.<sup>24</sup>

**Hypothesis 4a: There will be increased agreement between self and subordinates on second assessments in the category “STEPS to EMPOWERMENT” compared with first assessments.**

Z scores for this hypothesis are 3.45 and 2.54 at time one and time two, respectively. The Z score at time one is within the rejection region, unlike the Z score at time two. Therefore, there was sufficient evidence to reject the null hypothesis at time one, but not at time two. The decrease in difference in mean scores from time one to time two, .09 and .06 respectively, shows a convergence of self and subordinates' assessments at time two. Therefore, there is increased agreement between self and subordinates' assessments at time two, and support for this hypothesis. See Table 39 for mean scores for this hypothesis.

---

<sup>24</sup> See Appendix G for the results of Hypothesis 4 ANCOVA testing.

Table 39

Descriptive Statistics: Hypothesis 4a (Steps to Empowerment)

	N	Mean	Std. Deviation
STOE Self Time One	73	4.2952	.32619
STOE Subordinates Time One	499	4.2013	.73608
STOE Self Time Two	73	4.3438	.33218
STOE Subordinates Time Two	475	4.2794	.69076

**Hypothesis 4b: There will be increased agreement between self and subordinates on second assessments in the category “CREDIBILITY” compared with first assessments.**

Z scores for this hypothesis are 6.48 and 3.80 at time one and time two, respectively. Both are within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. The decrease in mean score differences from time one to time two, .20 and .11 respectively, shows a convergence of self and subordinates’ assessments at time two. Therefore, there is increased agreement between self and subordinates’ assessments at time two, and support for this hypothesis. See Table 40 for mean scores for this hypothesis.

Table 40

Descriptive Statistics: Hypothesis 4b (Credibility)

	N	Mean	Std. Deviation
CRED Self Time One	73	4.4514	.33980
CRED Subordinates Time One	499	4.2548	.77930
CRED Self Time Two	73	4.4292	.35710
CRED Subordinates Time Two	475	4.3196	.73289

**Hypothesis 4c: There will be increased agreement between self and subordinates on second assessments in the category “CHALLENGE and SUPPORT the ORGANIZATION” compared with first assessments.**

There was no support for this hypothesis. See Table 41 for mean scores.

Table 41

Descriptive Statistics: Hypothesis 4c (Challenge and Support the Organization)

	N	Mean	Std. Deviation
CSORG Self Time One	73	4.5137	.37035
CSORG Subordinates Time One	499	4.5107	.58436
CSORG Self Time Two	73	4.5114	.44125
CSORG Subordinates Time Two	475	4.5309	.58440

**Hypothesis 4d: There will be increased agreement between self and subordinates on second assessments in the category “BUILD TEAMS” compared with first assessments.**

There was no support for this hypothesis. See Table 42 for mean scores for this hypothesis.

Table 42

Descriptive Statistics: Hypothesis 4d (Build Teams)

	N	Mean	Std. Deviation
BT Self Time One	73	4.3562	.48218
BT Subordinates Time One	499	4.2806	.76766
BT Self Time Two	73	4.4247	.52156
BT Subordinates Time Two	475	4.3253	.73328

**Hypothesis 4e: There will be increased agreement between self and subordinates on second assessments in the category “DEVELOP INDIVIDUALS” compared with first assessments.**

There was a lack of support for this hypothesis. See Table 43 for mean scores.

Table 43

Descriptive Statistics: Hypothesis 4e (Develop Individuals)

	N	Mean	Std. Deviation
DI Self Time One	73	4.2661	.44308
DI Subordinates Time One	499	4.2897	.70335
DI Self Time Two	73	4.4012	.41626
DI Subordinates Time Two	475	4.3236	.69969

**Hypothesis 4f: There will be increased agreement between self and subordinates on second assessments in the category “QUALITY and TASK” compared with first assessments.**

There was a lack of support for this hypothesis. See Table 44 for mean scores for this hypothesis.

Table 44

Descriptive Statistics: Hypothesis 4f (Quality and Task)

	N	Mean	Std. Deviation
QT Self Time One	73	4.4315	.39224
QT Subordinates Time One	499	4.3583	.68353
QT Self Time Two	73	4.4425	.43745
QT Subordinates Time Two	475	4.4042	.65563

**Hypothesis 4g: There will be increased agreement between self and subordinates on second assessments in the category “CUSTOMER/CLIENT” compared with first assessments.**

Z scores for this hypothesis are 3.54 and 1.54 at time one and time two, respectively. The Z score at time one is within the rejection region and, therefore, there was sufficient evidence to reject the null hypothesis in this case, but not at time two. Therefore, there is a convergence of mean scores at time two as compared to time one. Thus, there is support for Hypothesis 4g. See Table 45 for mean scores for this hypothesis.

Table 45

Descriptive Statistics: Hypothesis 4g (Customer/Client)

	N	Mean	Std. Deviation
CC Self Time One	73	4.4132	.43842
CC Subordinates Time One	499	4.3069	.66752
CC Self Time Two	73	4.3425	.45550
CC Subordinates Time Two	475	4.3874	.59377

**Hypothesis 4h: There will be increased agreement between self and subordinates on second assessments in the category “VISION” compared with first assessments.**

There was a lack of support for this hypothesis. See Table 46 for mean scores for this hypothesis.

Table 46

Descriptive Statistics: Hypothesis 4h (Vision)

	N	Mean	Std. Deviation
V Self Time One	73	4.2849	.46776
V Subordinates Time One	499	4.2248	.77274
V Self Time Two	73	4.3096	.52709
V Subordinates Time Two	475	4.2825	.74648

**Hypothesis 4i: There will be increased agreement between self and subordinates on second assessments in the category “FOSTER CREATIVITY” compared with first assessments.**

Z scores for this hypothesis are 3.11 and 3.17 at time one and time two, respectively. Both are within the rejection. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. The decrease in mean score differences from time one to time two, .13 and .11 respectively, shows a convergence of self and subordinates’ assessments at time two. Therefore, there is increased agreement between self and subordinates’ assessments at time two, and there is support for this hypothesis. See Table 47 for mean scores for this hypothesis.



Table 47

## Descriptive Statistics: Hypothesis 4i (Foster Creativity)

	N	Mean	Std. Deviation
FC Self Time One	73	4.4178	.51884
Subordinates FC Time One	499	4.2866	.79186
FC Self Time Two	73	4.4315	.44733
Subordinates FC Time Two	475	4.3216	.74533

**Hypothesis 4j: There will be increased agreement between self and subordinates on second assessments in the category “MANAGE DIVERSITY” compared with first assessments.**

Z scores for this hypothesis are 10.25 and 7.06 at time one and time two, respectively. Both are within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. The decrease in mean score differences from time one to time two, .32 and .23 respectively, shows a convergence of self and subordinates’ assessments at time two. Therefore, there is increased agreement between self and subordinates’ assessments at time two, and support for this hypothesis. See Table 48 for mean scores for this hypothesis.

Table 48

Descriptive Statistics: Hypothesis 4j (Manage Diversity)

	N	Mean	Std. Deviation
MD Self Time One	73	4.7260	.41029
MD Subordinates Time One	499	4.4064	.73520
MD Self Time Two	73	4.6849	.42416
MD Subordinates Time Two	475	4.4522	.74288

**Hypothesis 4k: There will be increased agreement between self and subordinates on second assessments in the category “BUILD PUBLIC TRUST” compared with first assessments.**

Z scores for this hypothesis are 10.65 and 4.28 at time one and time two, respectively. Both are within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each of these cases. The decrease in mean scores from time one to time two, .19 and .10 respectively, shows a convergence of self and subordinates’ assessments at time two. Therefore, there is increased agreement between self and subordinates’ assessments at time two, and support for this hypothesis. See Table 49 for mean scores for this hypothesis.

Table 49

Descriptive Statistics: Hypothesis 4k (Build Public Trust)

	N	Mean	Std. Deviation
BPT Self Time One	73	4.7412	.32450
BPT Subordinates Time One	499	4.5536	.53126
BPT Self Time Two	73	4.6743	.40964
BPT Subordinates Time Two	475	4.5726	.53952

Finally, to further examine Hypothesis 4, ANCOVA was used on a global leadership scale by adding together all the data points in the 11 sub-categories and creating an overall mean score for subordinates. This was done for each of the cases at time one and time two. The between-subjects factors of 73 mean self scores were then compared with 70 mean subordinate scores. The dependent variable was the mean overall scale at time two and the covariate was the mean overall scale at time one. Results showed a non-significant alpha level of .31. It suggests a lack of support for this hypothesis from a global perspective, which is also confirmed by the findings of five of the sub-hypotheses and the t tests. Table 50 reflects the between-subjects effects of this test.

Table 50

## ANCOVA: Hypothesis 4 (Overall Measurement)

## Tests of Between-Subjects Effects

Dependent Variable: ALL2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.086 <sup>a</sup>	2	2.543	20.380	.000
Intercept	4.110	1	4.110	32.938	.000
ALL1	4.665	1	4.665	37.393	.000
RESPTYPE	.132	1	.132	1.057	.306
Error	17.468	140	.125		
Total	2770.830	143			
Corrected Total	22.553	142			

a. R Squared = .225 (Adjusted R Squared = .214)

Because there was a lack of support for five sub-hypotheses (4c, 4d, 4e, 4f, and 4h), and t tests showed a non-significant results in comparing time one difference of means to time two difference of means in all sub-categories, there was also a lack of support for Hypothesis 4. However, as in the case of Hypotheses 2 and 3, by testing these sub-hypotheses valuable information regarding the subscales as measuring different leadership categories of the LCCIS was developed. The implications will be discussed in Chapter 5.

#### *Hypothesis 5*

To add further depth and richness to this study, agreement between subordinates and peers on first and second assessments were compared. While this issue is different than the majority focus in this research, which is on self-assessments versus others' assessments, an additional level of understanding regarding multi-rater feedback was developed. The following hypothesis is offered which addresses the sixth and final research question of this study:

6. Are there differences in agreement between peer and subordinate ratings at time one and time two?

**Hypothesis 5: Peers' ratings will be more similar to self-ratings on first and second assessments than will subordinates' ratings.**

This hypothesis complements earlier hypotheses, which suggested that assessed individuals would rate themselves higher than peers and subordinates on first

assessments, and that there would be convergence on second assessments. Hypothesis 5 suggests that peers' ratings would align with self-ratings more than subordinate ratings.

In order to test this hypothesis, first a mean of all 73 cases of self scores at time one was determined for comparison with an overall mean score of peers at time one. A mean of all 73 cases of self scores at time two was then calculated for comparison to an overall mean score of peers at time two (see Table 51). Z scores were then computed for comparison of means at time one and comparison of means at time two in order to test the null hypothesis that the means in either instance did not differ. The observed value of  $Z = -6.48$  at time one, and the observed value of  $Z = 4.47$  at time two were within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each case. Consequently, the results indicate a statistical difference between perceptual agreement measures of self versus peers at time one and at time two.

Further, a mean of all 73 cases of self scores at time one was determined for comparison with an overall mean score of subordinates at time one. A mean of all 73 cases of self-scores at time two was calculated for comparison to an overall mean score of subordinates at time two (see Table 51). The observed value of  $Z = 5.52$  at time one, and the observed value of  $Z = 3.40$  at time two were within the rejection region. Therefore, there was sufficient evidence to reject the null hypothesis in each case. Consequently, the results indicate a statistical difference between perceptual agreement measures of self versus subordinates at time one and time two.

While these Z score computations were found to be significant, they only reveal part of the story. The patterns in mean differences are of interest, as self mean scores in each instance (time one and time two) are less than peer mean scores, and self mean scores in each instance are greater than subordinate mean scores. However, self and peer mean scores at time one are in greater agreement (.09 difference) compared with self and subordinate mean scores (.12 difference). At time two, self and subordinate mean scores are in greater agreement (.07 difference) compared with self and peer mean scores (.09 difference). This does not lend support for Hypothesis 5. Further, t tests in all sub-categories of peers versus self showed a non-significant results in comparing time one difference of means to time two difference of means. These implications will be discussed further in Chapter 5.

Table 51

Descriptive Statistics: Hypothesis 5 (Overall Ratings)

	N	Mean	Std. Deviation
ALL Self	73	4.4324	.30176
Time One Peers	442	4.5247	.44599
Time One Subordinates	499	4.3124	.65044
ALL Self	73	4.4403	.33565
Time Two Peers	386	4.5278	.53257
Time Two Subordinates	475	4.3663	.61960

To further address Research Question 6 time one peer and subordinate scores were directly compared with one another, as were time two scores. The results showed statistically significant difference in each instance. The time one difference of means was .2123 and the time two difference of means was .1615. These results indicate a difference in agreement between peer and subordinate ratings at time one and time two.

ANCOVA testing was also done to test between-subjects effects of peers and subordinates as an overall scale and on each of the individual subscales corresponding to the categories offered in Hypotheses 2, 3, and 4. Although not the focal point of this research, this testing adds even greater depth, understanding and robustness to this research.

The results of the ANCOVA showed non-significant alpha levels in every case except the VISION and FOSTER CREATIVITY sub-categories. It is noted however, the overall results indicate marginal significance effects. The alpha level for the twelve tests of peers to subordinates range from p equal to or < .04 to .18.<sup>25</sup> The ANCOVA test on a global or overall leadership scale comparing between-subjects factors of 73 mean peer

<sup>25</sup> See Appendix H for the results of Hypothesis 5 ANCOVA testing

scores with 73 mean subordinate scores is typical and a good reflection of these twelve tests with an overall significance level of .06. Results of this test are depicted in Table 52.

Table 52  
ANCOVA: Hypothesis 5 (Overall Measurement)

**Tests of Between-Subjects Effects**

Dependent Variable: ALL2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.035 <sup>a</sup>	2	3.517	28.424	.000
Intercept	4.235	1	4.235	34.224	.000
ALL1	5.523	1	5.523	44.632	.000
RESPTYPE	.445	1	.445	3.592	.060
Error	17.324	140	.124		
Total	2835.240	143			
Corrected Total	24.359	142			

a. R Squared = .289 (Adjusted R Squared = .279)

Finally, in order to contribute to a discussion of these differences from an overall perspective, Table 53 is offered as a summary of numeric differences in means. The corresponding hypotheses are reflected within the table.

Table 53  
 Numeric Differences in Means

	Self / Others	Self / Peers	Self / Subordinates	Peers / Subordinates <sup>1</sup>
<b>ALL</b>				
T1 <sup>2</sup>	.0260 (Hyp 1)	<b>-.0795**</b> <sup>3</sup> (Hyp 5)	<b>.1151**</b> (Hyp 5)	1945 (Hyp 5)
T2 <sup>4</sup>	.0192 (Hyp 1)	<b>-.0973**</b> (Hyp 5)	<b>.1096**</b> (Hyp 5)	.2068 (Hyp 5)
T2 – T1	-.0068 (Hyp 1)	-.0178 (Hyp 5)	-.0055 (Hyp 5)	.0123 (Hyp 5)
<b>STOE</b>				
T1	.0027 (Hyp 2a)	<b>-.1068**</b> (Hyp 3a)	<b>.0890**</b> (Hyp 4a)	.1959 (Hyp 5)
T2	.0000 (Hyp 2a)	<b>-.0986**</b> (Hyp 3a)	.1055 (Hyp 4a)	.2041 (Hyp 5)
T2 – T1	-.0027 (Hyp 2a)	.0082 (Hyp 3a)	.0164 (Hyp 4a)	.0082 (Hyp 5)
<b>CRED</b>				
T1	<b>.0849**</b> (Hyp 2b)	<b>-.0438**</b> (Hyp 3b)	<b>.1945**</b> (Hyp 4b)	.2384 (Hyp 5)
T2	.0507 (Hyp 2b)	<b>-.0712**</b> (Hyp 3b)	<b>.1521**</b> (Hyp 4b)	.2233 (Hyp 5)
T2 – T1	-.0342 (Hyp 2b)	-.0274 (Hyp 3b)	-.0425 (Hyp 4b)	-.0151 (Hyp 5)
<b>CSO</b>				
T1	<b>-.0740**</b> (Hyp 2c)	<b>-.1534**</b> (Hyp 3c)	.0041 (Hyp 4c)	.1575 (Hyp 5)
T2	-.0603 (Hyp 2c)	<b>-.1479**</b> (Hyp 3c)	.0082 (Hyp 4c)	.1562 (Hyp 5)
T2 – T1	.0137 (Hyp 2c)	.0055 (Hyp 3c)	.0041 (Hyp 4c)	-.0014 (Hyp 5)
<b>BT</b>				
T1	-.0356 (Hyp 2d)	<b>-.1589**</b> (Hyp 3d)	.0753 (Hyp 4d)	.2342 (Hyp 5)
T2	.0370 (Hyp 2d)	-.0849 (Hyp 3d)	.1356 (Hyp 4d)	.2205 (Hyp 5)
T2 – T1	.0726 (Hyp 2d)	.0740 (Hyp 3d)	.0603 (Hyp 4d)	-.0137 (Hyp 5)
<b>DI</b>				
T1	<b>-.0918**</b> (Hyp 2e)	<b>-.1781**</b> (Hyp 3e)	-.0247 (Hyp 4e)	.1534 (Hyp 5)
T2	.0247 (Hyp 2e)	<b>-.0767**</b> (Hyp 3e)	<b>.1082**</b> (Hyp 4e)	.1849 (Hyp 5)
T2 – T1	.1164 (Hyp 2e)	.1014 (Hyp 3e)	.1329 (Hyp 4e)	.0315 (Hyp 5)
<b>QT</b>				
T1	-.0027 (Hyp 2f)	<b>-.1000**</b> (Hyp 3f)	.0740 (Hyp 4f)	.1740 (Hyp 5)
T2	-.0151 (Hyp 2f)	<b>-.1205**</b> (Hyp 3f)	.0808 (Hyp 4f)	.2014 (Hyp 5)
T2 – T1	-.0123 (Hyp 2f)	-.0205 (Hyp 3f)	.0068 (Hyp 4f)	.0274 (Hyp 5)
<b>CC</b>				
T1	.0438 (Hyp 2g)	-.0288 (Hyp 3g)	<b>.0932**</b> (Hyp 4g)	.1219 (Hyp 5)
T2	<b>-.0740**</b> (Hyp 2g)	<b>-.1411**</b> (Hyp 3g) <sup>5</sup>	-.0219 (Hyp 4g)	.1192 (Hyp 5)
T2 – T1	-.1178 (Hyp 2g)	-.1123 (Hyp 3g)	-.1151 (Hyp 4g)	-.0027 (Hyp 5)
<b>V</b>				
T1	-.0219 (Hyp 2h)	<b>-.1329**</b> (Hyp 3h)	.0575 (Hyp 4h)	.1904 (Hyp 5)
T2	-.0288 (Hyp 2h)	<b>-.1493**</b> (Hyp 3h)	.0740 (Hyp 4h)	<u>.2233 (Hyp 5)</u>
T2 – T1	-.0068 (Hyp 2h)	-.0164 (Hyp 3h)	.0164 (Hyp 4h)	.0329 (Hyp 5)
<b>FC</b>				
T1	.0836 (Hyp 2i)	.0329 (Hyp 3i)	<b>.1329**</b> (Hyp 4i)	.1000 (Hyp 5)
T2	.0740 (Hyp 2i)	-.0219 (Hyp 3i)	<b>.1548**</b> (Hyp 4i)	<u>.1767 (Hyp 5)</u>
T2 – T1	-.0096 (Hyp 2i)	-.0548 (Hyp 3i)	.0219 (Hyp 4i)	.0767 (Hyp 5)
<b>MD</b>				
T1	<b>.2096**</b> (Hyp 2j)	<b>.0740**</b> (Hyp 3j)	<b>.3151**</b> (Hyp 4j)	.2411 (Hyp 5)
T2	<b>.1493**</b> (Hyp 2j)	.0137 (Hyp 3j)	<b>.2466**</b> (Hyp 4j)	.2329 (Hyp 5)
T2 – T1	-.0603 (Hyp 2j)	-.0603 (Hyp 3j)	-.0685 (Hyp 4j)	-.0082 (Hyp 5)

<sup>1</sup> Significance testing was not reported here on any peers versus subordinate ratings as it was outside the scope of Hypothesis 5 as offered. This column is presented simply to add more depth to Table 53.

<sup>2</sup> Time One

<sup>3</sup> Bold type and\*\* indicates p < or equal to .05.

<sup>4</sup> Time Two

<sup>5</sup> Underline indicates ANCOVA results of p < or equal to .05.



**Table 53 (Cont.)**

<b>BPT</b>				
T1	<b>.1027**</b> (Hyp 2k)	-.0096 (Hyp 3k)	<b>.1904**</b> (Hyp 4k)	.2000 (Hyp 5)
T2	.0342 (Hyp 2k)	<b>-.0808**</b> (Hyp 3k)	<b>.1288**</b> (Hyp 4k)	.2096 (Hyp 5)
T2 – T1	-.0685 (Hyp 2k)	-.0712 (Hyp 3k)	-.0616 (Hyp 4k)	.0096 (Hyp 5)

### Conclusion

This chapter discussed the analysis that developed from this study. It began with a summary of the purpose of the research and was followed by an instrument summary and data analysis. The data analysis addressed each of the research questions as detailed through five primary hypotheses and 33 sub-hypotheses. A significant number of findings were generated as result of this effort. The next chapter will discuss these findings and their implications. It will conclude with the study's limitations and recommendations for further study and practice.

## CHAPTER 5

### Summary and Implications

#### Introduction

This chapter sets forth a summary of the study presented here. It also presents the conclusions of this research. It begins by providing an overview of the findings and is followed by a discussion regarding the outcomes and implications of the study. Finally, this chapter ends with the limitations and recommendations for further research and practice.

#### Overview

According to Martineau (1998), using any type of instrumentation, multi-rater or not, to measure the success of development requires it to be interpreted within the context of the system. This research tested a part of the multi-rater system, which is in place in a public sector law enforcement setting, by examining self versus subordinates and peer ratings at two times. Therefore, the interpretation of the results must be considered within the context of the organization, its attendant idiosyncrasies, and within the limitations of the present study.

Specifically, this study examined the perceptions of individuals with regard to leadership dimensions in the context of the FBI work setting. There is evidence that no statistically significant differences exist in the perceptions of leadership between self and others, as an overall measurement on either a first assessment or a second assessment. Additionally, it is noted that in every hypotheses and sub-hypotheses t tests examining difference of means between first and second assessments yielded non-significant alpha levels. As such, the following discussion is a relative one. Nonetheless, when analyzing subscales or specific dimensions of leadership behavior, the results are both interesting and varied. Additionally, generally supervisors rated themselves lower than their peers and higher than subordinates. These findings are not altogether consistent with the literature and indicate that perhaps some contextual interpretations may be associated with the participants of this research. That is, all the individuals used in this study are career law enforcement professionals who underwent a very thorough background

investigation prior to entry in the FBI. These individuals may have a homogeneous worldview including their perspective of leadership issues, which may be reflected in the findings. Further, the results must be interpreted within a fairly narrow range of mean scores for most of the categories, basically between 4.2 and 4.7 on a five point Likert Scale. Thus, every mean comparison becomes a relative one. Additionally, there were only 73 individuals who were the participants of this research due to limitations of those available at the time of this study.

Finally, within the dimensions of MANAGE DIVERSITY and BUILD PUBLIC TRUST, two categories that were part of the testing instrument, there were interesting findings that lent a unique perspective to this particular research. Specifically, supervisors rated themselves more highly on these particular dimensions than others (peers and subordinates) as indicated by difference in mean averages at both time one and time two. Generally, the findings also showed that peers and subordinates rated these leaders more highly in these dimensions than they rated them on other dimensions as indicated by mean averages. As with the overall findings, these results may indicate that there are contextual interpretations that have to do with a prescreened law enforcement population, and how individuals in this population perceive themselves. Further, it becomes clear that these are critical dimensions to the culture of the FBI.

## Discussion

One of the principal findings of multi-source research on the level of agreement between self-raters and other raters is that there is generally little agreement between self and others (Atwater & Yammarino, 1992). Furthermore, there is a propensity for individuals to overrate themselves as compared to others (Thornton, 1980; Harris & Schaubroeck, 1988). The following discussion provides a different perspective from these previous findings as was determined through the outcome of the Research Questions and hypotheses offered. It does so by providing some view into multi-rater feedback by examining the area of perceptual agreement among multi-rater feedback sources in an effort to assess these previous findings. It adds to the literature by offering a perspective about perceptual agreement between multi-rater feedback sources as part of a case study in a large public sector law enforcement organization – the FBI. Prior to this

work there was not much research using multi-rater feedback in the law enforcement domain. As such, the findings offered here are a piece of scholarship that helps to shed light on both the uniqueness of a law enforcement culture and some of the attendant idiosyncrasies associated with it.

Significantly, the FBI is a fairly insulated culture and the findings of this research may be a reflection of adherence to mores associated with it. What was being measured could have been this adherence as much as leadership. Additionally, all FBI employees undergo a very rigorous and thorough background investigation prior to their entry into the organization. Through this process alone many “would be” leaders and raters were not involved in this study. These facts highlight two things. One is the importance of examining the context of this study and understanding where it fits within the broad scope of multi-rater feedback research. The second is being aware that this research is but one leadership intervention in use by the organization, and has a contextual place within the FBI as well.

This research opened the door to what has traditionally been a fairly insulated work setting. The results illuminate an area of multi-rater feedback that has not been explored extensively, namely its use in a law enforcement environment. The value added is in exposing the world to a part of the FBI that has not had scrutiny and expanding the understanding of the multi-rater feedback process in a place heretofore where there is little or no documentation on the subject.

Further, in more closely assessing the overall value of what was learned in this instance, the research conducted highlights a specific part of the FBI leadership programs, EDI I and EDI II. However, the multi-rater instrument used for this study is not the only measurement of leadership in use by the organization. Annual performance evaluations, informal mentoring, temporary additional duties, and a variety of other training opportunities all serve to augment the multi-rater feedback experience. As such its significance and the analysis must be considered as but one piece of the overall human development process in the organization. Further, the results must be assessed with the understanding these other experiences may have affected the study’s outcome. All these pieces of the “FBI experience” add to a rich and highly developed insular familiarity to those within. The instrument itself is a part of this familiarity. After more than 15 years

of use, it is as much a part of the culture of the organization as many of the other artifacts within it. The fact that part of this instrument was designed by the organization itself and measures the concepts MANAGE DIVERSITY and BUILD PUBLIC TRUST must also be kept in mind. All these aspects of this research are both benefits that help to advance scholarship in this area and are also contextual restrictions.

Regarding specifics, the instrument used in this study contained 11 subscales that could be assessed. The present research examined these sub-scales using six research questions and five hypotheses. Three of the hypotheses contained 11 sub-hypotheses corresponding to each of the subscales of the LCCIS. Therefore, distinctions could be addressed if found.

The study evidence showed this: in many dimensions there was more alignment during the second rating than on the first as predicted in the hypotheses. However, when comparing differences in the amount of convergence at time two compared with time one there was no statistical difference for all hypotheses.

In the main it was found that the hypotheses were not supported, however when moving from the overarching findings into the nuances, there were a lot of interesting developments not altogether expected. Clearly, this research showed evidence that those rated (aspiring executives in the FBI) see themselves differently from their peers and differently from their subordinates, and on assorted leadership dimensions. These breakdowns will be addressed below.

#### *Research Question 1*

This section examines the findings by focusing on the six research questions that were the concentration of this study. The discussion continues by examining the findings surrounding the comparison of self with “others,” the combination of peers and subordinates. It does so by addressing the first research question, which is:

1. Is there a greater perceptual agreement between others (peers and subordinates) and self on the second multi-rater assessment than the initial assessment?

As an overall measurement comparing self to others (Hypothesis 1), there was no statistical difference found between the ratings of self versus others at time one or self versus others at time two. By comparing others at time one (as an overall measurement) to self at time two there was no statistically significant difference, either. T tests

comparing difference of means between first and second assessments were also consistent with these results. Further, ANCOVA showed no statistical difference at time two when controlling for time one scores. These results all indicate that leaders fundamentally did not change their overall view of their leadership ability from the first time they received the assessment to the second time they received it as measured by the LCCIS. Further, while not setting out to specifically examine this outcome, Hypothesis 1 testing also determined that the means of self versus others at both time one and two were almost identical (4.4 on a 5 point Likert scale). The implications are that as an overall measure of leadership, the way leaders in the FBI rate themselves compared with how others (subordinates and peers) rate them is basically the same.

However, while the Z score computations for difference in mean comparisons at time one and time two were not found to be significant, the patterns in mean differences were of interest and the difference in means (.0203 at time one > .0016 at time two) suggests a very small amount of support for Hypothesis 1. While statistically insignificant relative to the size of these differences, it is consistent with the hypothesis. These implications suggest that there is some, though very limited, support for the idea that agreement increased between self and others on second assessments in this overall view of leadership compared with first assessments.

Additionally, although a negative mean difference was found between others at time one and self at time two (- .0282), which was not statistically significant (as measured by a Z score), the result indicates more leadership on the second assessment. That is, the recipients of the feedback rated themselves more highly at time two than others did at time one. This result, combined with a smaller mean difference between self and others at time two versus time one, may indicate better leadership and the desired overall outcome. Ideally, leaders will learn about how others perceive them from feedback in the first instance, make adjustments to their behavior in the direction of improvement, and in their second assessment align their ratings more closely with those who rate them. The results for Research Question 1 shows no statistically significant support in the direction of this outcome, but some patterns that support using multi-rater feedback as a productive tool for leadership improvement in the FBI.

Because the findings here show no statistically significant support in the direction of Research Question 1 (Hypothesis 1), they are inconsistent with Atwater et al.'s (2000) research that indicated leaders rating themselves and receiving feedback from others twice through multi-rater assessment brought them more into alignment with their subordinates' ratings. These authors suggest that personal growth can take place if individuals who overrate themselves on a first assessment align their ratings with others on second assessments. Their findings were not confirmed in this study.

Additionally, according to Brutus, et al. (1998) and London and Tornow (1998), the desired behavioral outcome of using 360-degree feedback is to improve the leadership capacity of the individuals being assessed. Here the results indicate that those outcomes were not confirmed except through non-statistical support. These findings are of great interest. Clearly, the organization being studied has devoted a tremendous amount of energy, effort and time to the multi-rater instrument process. Yet, the results suggest there are no statistical significant differences found between self and others' perceptions of the leaders even after they have benefited from receiving the results of the initial feedback. What are the implications for the use of multi-rater feedback and leadership?

It would appear that those individuals who were used in this study are capable leaders who begin with a strong self-perception of their abilities. The record shows high results on a Likert scale for self and others at time one and time two. Evidently, there is a healthy self concept about leadership ability in the FBI, which is confirmed by others' ratings.

It also seems that in the main these leaders did not appear to benefit significantly by the impact of the multi-rater process or did not perceive a significant need to change based on the results learned at time one. In some measure it suggests that if learning did take place through critical reflection (Brookfield 1990) at time one it did not result in changed behavior as indicated by the LCCIS time two results. The implications may be that a more thorough system that includes individual coaching plans could help in instances where those rated find value in feedback but do not know how to translate it into an actionable agenda. An individual coaching plan with an action component can be used as an accompaniment to the multi-rater feedback process to make it more effective. In this study those assessed were left to their own device to use the information obtained

from these instruments. A more structured individual development plan combined with multi-rater feedback may have changed the results of this study.

These results may also point to the fact that those leaders used in this study were a fairly homogeneous group with similar attitudes about personal and professional development. Here they were tested twice to determine their leadership ability and were provided some training for improvement. Yet, after both iterations of the training they were sent back into a work environment where they were busy attending to operational issues with little chance for reflection on the training provided or the results of the feedback. If the work environment to which they returned did not attach much importance to the training and support it with appropriate follow-on assignments it had a limited effect for changing behavior. While this was not independently tested it suggests implications for goal setting used in conjunction with multi-rater feedback discussed in the “Recommendations for Further Research” section below.

Finally, these overall results may indicate that those used in this study represent only a percentage of the leaders in the organization and perhaps the better ones. These individuals volunteered and were selected for positions of greater responsibility in the FBI. This was evident by the fact that they were in the respective training where the LCCIS testing was done. The results show high self ratings which were confirmed by others’ ratings. No changes at time two may represent a laissez-faire attitude by those assessed toward the need for critical reflection or change in the way they were leading. While the results measured here represent a standard of good leadership or even very good leadership it does not go to the heart of reaching the full potential of the leadership capacity within the organization. In some respects it further muddies the water regarding multi-rater feedback when leadership levels appear to be fairly high. It lends credence to the conclusion of Atwater et al. (2000), “[that] the evidence regarding the impact of feedback on self-ratings is not altogether clear” (p. 278).

### *Research Question 2*

While the first question considered an overall measure, it may not convey the entire story. Further discussion will clarify how the 11 subscales of the LCCIS contribute to this research.



In order to more carefully examine the subcomponents of the LCCIS, the second question dealt with the nature of change in agreement between self-ratings and others' (peer and subordinate) ratings on second assessments as compared to first assessments. It is:

2. What are the dimensions (sub-categories of the LCCIS) of any differences found?

As the discussion moves beyond the overall scale used in the first research question, the findings are mixed for self versus others as the various subscales are considered. Hypothesis 2 addressed this issue.

Because there was a lack of support for six out of 11 sub-hypotheses (corresponding to each of the 11 dimensions of the LCCIS) of Hypothesis 2 (2a, 2d, 2f, 2g, 2h, and 2i), and t tests showed a non-significant results in comparing time one difference of means to time two difference of means in all sub-categories, there was also a lack of support for Hypothesis 2. This hypothesis suggested leaders in the FBI would rate their leadership more highly compared with others (peers and subordinates together) on first assessments compared with second assessments in each subcategory. This hypothesis also suggested that leaders would more closely align their self assessments with others on second evaluations than on first. However, the mixed findings suggest that when grouping peers and subordinates together into "others," there is little consistency in the results.

In answering Research Question 2 the results indicated the following. The five leadership dimensions in which there was evidence of some (non-statistically significant) support for the sub-hypotheses of Hypothesis 2 were CREDIBILITY, CHALLENGE and SUPPORT the ORGANIZATION, DEVELOP INDIVIDUALS, MANAGE DIVERSITY, and BUILD PUBLIC TRUST. By contrast, the six dimensions where there was a lack of support for the sub-hypotheses were STEPS to EMPOWERMENT, BUILD TEAMS, QUALITY and TASK, CUSTOMER/CLIENT, VISION, and FOSTER CREATIVITY. There seems to be no indication of an overall pattern regarding specific dimensions.

The five dimensions where there was non-statistically significant evidence of support for Hypothesis 2 indicate there is a change in perceptions about the leaders'

abilities in the desired direction of more alignment at time two than time one. These results suggest that in roughly half (five of 11) of the dimensions evaluated in this study the outcomes indicate the desired effect. Further, the breakdown of the subcategories offer something of an understanding and indicate there are some differences in the sub-scales, which balance out when all measures are combined into an overall scale (Hypothesis 1).

ANCOVA testing shed some light on Hypothesis 2. In every subcategory for this hypothesis the ANCOVA results indicated non-significant alpha levels. Further, an ANCOVA test on a global or overall leadership scale comparing between-subjects factors of 73 mean self scores with 73 mean other scores also showed non-significant alpha levels. It appears as if the initial response (time one) type makes little difference in the final response (time two) when time one scores are controlled for.

### *Research Question 3*

However, the fact that as different groups peers and subordinates rated supervisors consistently across leadership dimensions was evident more often than not in this study. This was regularly confirmed in almost all the sub-hypotheses dealing with self versus peers (Hypotheses 3a - k), in which difference in mean self scores were lower than difference in mean peer scores. It also was regularly confirmed in almost all the sub-hypotheses dealing with self versus subordinates (Hypotheses 4a – k), in which difference in mean self scores were greater than difference in mean subordinate scores. Research Question 3 dealt with this issue. It is:

3. Are there differences in agreement between peer and self-ratings and subordinate and self-ratings at time two and time one?

As indicated, the results generally demonstrated these differences. In nine out of 11 sub-hypotheses dealing with self mean scores compared with peer mean scores (Hypothesis 3), the supervisors underrated themselves compared with peers at time one and time two. This was reflected by negative mean differences in each instance, regardless of whether the differences were significant or not (as measured by a Z score). The only categories in which they did not were FOSTER CREATIVITY and MANAGE DIVERSITY. Further, in the same nine out of 11 categories, statistically significant differences were found in time two mean self scores compared with time one mean peer

scores. In each of these instances this confirmed that mean difference self scores at time two were lower than the mean peer scores at time one.

What these results imply are a self-perception of less leadership for those rated at time two compared with the assessment by peers at time one. The inference seems to be that peers in the FBI work setting perceive their equals as more qualified than those rated see themselves. This remains true on second assessments even after those rated had the benefit of receiving and digesting the initial feedback. These findings are interesting and may point to how much leaders do not know, possibly indicating opportunity for continued development. Certainly it illuminates a unique perspective in the area of multi-rater feedback.

Previous findings in this area by Landy and Farr (1983) indicated that multi-rater feedback from peers was found to be valid and reliable. The findings in this research show enough evidence to indicate that peers either have an inflated sense of one another's leadership ability or give themselves the benefit of the doubt when they are asked to rate each other. This may be caused by the fact that they have a unique understanding of each other's roles and responsibilities. It may also indicate empathy for the struggles of the "leadership" position. Further, it may evidence that there is a higher order cognitive skill process associated with leadership that is only understood by those who have shouldered a similar responsibility. Too, it may be indicative of a maturing process that comes with important assignments associated with ever-greater leadership positions. Whatever the case, the perception is not shared by the subordinates in this research.

Generally the opposite was true of subordinates' ratings of their supervisors. In eight out of 11 sub-hypotheses dealing with self mean scores compared with subordinate mean scores (Hypothesis 4), the supervisors overrated themselves compared with subordinates at time one and time two. This was reflected by positive mean differences in each instance, regardless of whether the differences were statistically significant or not (as measured by a Z score). The only categories in which the supervisors did not overrate themselves at both times were CHALLENGE and SUPPORT the ORGANIZATION, DEVELOP INDIVIDUALS, and CUSTOMER/CLIENT. Yet, in each of these instances at either time one or time two, they overrated themselves at one time or the other compared with their subordinates.

Further, in eight out of 11 categories, statistically significant differences were found in time two mean self scores compared with time one mean subordinate scores. The only categories where this was not the case were CHALLENGE and SUPPORT the ORGANIZATION, QUALITY and TASK, and CUSTOMER/CLIENT. The positive instances confirmed that mean self scores at time two were higher than mean subordinate scores at time one. It implies a self-perception of more leadership for those rated at time two compared with the assessment by subordinates at time one. The implications are that in general subordinates in the FBI work setting see their supervisors as having less leadership ability than those rated see themselves as having. This general finding remains true on second assessments even after those rated have the benefit of receiving and digesting the initial feedback results.

Similar to the peer rating findings, these findings are interesting. Unlike the peer rating findings, these results are more consistent with previous research on multi-rater assessment where self raters tend to rate themselves higher than others.

As Borman (1991) indicates, the advantages of multi-rater feedback are that each of the raters, whether they are direct reports, peers, or the supervisor, may provide germane but slightly different information. Here we find that subordinates clearly see their leaders differently than the leaders' peers do or than the leaders see themselves. By and large, subordinates' perceptions are lower than the peers. Yet, when combined (Hypothesis 1), they cancel each other out.

#### *Research Questions 4 & 5*

The findings in this research show evidence that subordinates either have a depressed sense of their supervisors' leadership ability or do not give them the benefit of the doubt. This may be caused by the fact that they do not have an understanding of their bosses' roles and responsibilities. It may also indicate a lack of empathy for the struggles of the supervisory position. Further, it may evidence that there is a lack of insight and discernment associated with leadership when someone has not shouldered a similar responsibility. Or it may be indicative of a lack of maturity that can only come with ever-greater leadership positions. Whatever the case, the perceptions of subordinates are clearly different than those of the peers in this study.

Research Questions 4 and 5 get at the heart of these differences. They are:

4. What are the dimensions of any differences found in peer and self-ratings?
5. What are the dimensions of any differences found in subordinate and self-ratings?

Hypothesis 3 addressed the fourth research question and separated “others” (from the first hypothesis) into the portion made up of peers. Hypothesis 4 addressed the fifth research question, and separated “others” into the portion made up subordinates.

Results indicated a lack of support for Hypothesis 3 (the hypothesis dealing with peers) because there was a lack of support for four of its sub-hypotheses (3b, 3g, 3i, and 3k) and t tests showed non-significant results in comparing time one difference of means to time two difference of means in all sub-categories. Results also indicated a lack of support for Hypothesis 4 (the hypothesis dealing with subordinates) because there was a lack of support for five of its sub-hypotheses (4c, 4d, 4e, 4f, and 4h) and t tests showed non-significant results in comparing time one difference of means to time two difference of means in all sub-categories.

The seven leadership dimensions in which there was evidence of some non-statistically significant support for the sub-hypotheses of Hypothesis 3 (that is, evidence of more alignment at time two than time one) were STEPS to EMPOWERMENT, CHALLENGE and SUPPORT the ORGANIZATION, BUILD TEAMS, DEVELOP INDIVIDUALS, QUALITY and TASK, VISION, and MANAGE DIVERSITY. By contrast the four dimensions where there was a lack of support for the sub-hypotheses were CREDIBILITY, CUSTOMER/CLIENT, FOSTER CREATIVITY, and BUILD PUBLIC TRUST.

The six dimensions in which there was evidence of some non-statistically significant support for the sub-hypotheses of Hypothesis 4 (that is, evidence of more alignment at time two than time one) were STEPS to EMPOWERMENT, CREDIBILITY, CUSTOMER/CLIENT, FOSTER CREATIVITY, MANAGE DIVERSITY, and BUILD PUBLIC TRUST. By contrast the five dimensions where there was a lack of support for the sub-hypotheses were CHALLENGE and SUPPORT the ORGANIZATION, BUILD TEAMS, DEVELOP INDIVIDUALS, QUALITY and TASK, and VISION.

In each of these cases (Hypothesis 3 and Hypothesis 4) there seems to be little indication of an overall pattern regarding specific dimensions. Similar to Research Question 2, the findings from the specific sub-hypotheses imply a randomness to the results, but indicate that on some leadership dimensions there is the desired outcome of more alignment at time two than time one.

The two dimensions where there was a consistency of results between Hypotheses 3a - k and 4a - k were STEPS to EMPOWERMENT and MANAGE DIVERSITY. In these two dimensions there is evidence that with both peers and subordinates, there is a change in perceptions about the leaders' abilities in the desired direction of more alignment at time two than time one though this was not confirmed by the difference of means t test. These results indicate there is a change in perceptions about the leaders' abilities in the desired direction of more alignment at time two than time one though not statistically so.

ANCOVA testing helped shed some light on Hypotheses 3 and 4. In every case for both these hypotheses except the CUSTOMER/CLIENT sub-category for Hypothesis 3 (self versus peers) the results showed non-significant alpha levels. Additionally, ANCOVA tests on global or overall leadership scales comparing between-subjects factors of mean self scores with mean peer scores and mean self scores with mean subordinate scores showed non-statistical alpha levels. It appears as if the initial response (time one) type makes little difference in the final response (time two) when time one scores are controlled for with either of these groups.

However as was also discovered in this research, by examining the results beyond the hypotheses' outcomes in most of these dimensions (regardless of statistical significance), peers overrated FBI leaders and subordinates underrated them. This was as compared to the leaders' self ratings. This may be the most important finding of this study and suggests that in the FBI work setting the perception of leaders is dependent upon the particular leadership status of those rating the leader.

#### *Research Question 6*

Finally, to add further depth and richness to this study, agreement between subordinates and peers on first and second assessments were compared. While this issue is different than the majority of the focus of this research, which is on self-assessments

versus others' assessments, an additional level of understanding regarding multi-rater feedback was developed. The following research question was offered which addresses this issue:

6. Are there differences in agreement between peer and subordinate ratings at time one and time two?

This question complements earlier questions, and asks directly what were the differences in agreement between peer and subordinate ratings at different times. This question was addressed by Hypothesis 5, which suggested that peers' ratings would align with self-ratings more than subordinate ratings at both time one and time two. This question was also addressed by directly comparing peer and subordinate mean differences at time one and time two.

The overall results showed a lack of support for Hypothesis 5. By examining the patterns in mean differences in each instance (time one and time two) the results show self and peer mean scores at time one in greater agreement (.09 difference) compared with self and subordinate mean scores (.12 difference). At time two however, self and subordinate mean scores were in greater agreement (.07 difference) compared with self and peer mean scores (.09 difference).

However, it was also learned that when peer and subordinate mean scores at time one and two were directly compared, they were found to be statistically significant different (as measured by a Z score). These findings reinforce what was discussed earlier regarding peer and subordinate perceptions. Specifically, peers and subordinates in the FBI work setting perceive their leaders differently.

These same results were also indirectly confirmed through the ANCOVA testing for Question 6. Particularly, ANCOVA was done to test between-subjects effects of peers and subordinates as an overall scale and on each of the individual subscales (corresponding to the same categories offered in Hypotheses 2, 3, and 4). The results of the ANCOVA for peers and subordinates showed non-significant alpha levels in every dimension except the VISION and FOSTER CREATIVITY sub-categories. It is noted however, and this an important point, the overall results indicate marginal significance effects. The significance level for the twelve tests of peers to subordinates range from p equal to or < .04 to .18. The ANCOVA test on a global or overall leadership scale

comparing between-subjects factors of mean peer scores with mean subordinate scores is typical and a good reflection of these twelve tests with an overall significance level of .06. These results (though not all statistically significant) supplement the Z score testing.

The conclusion drawn from all these outcomes from Question 6 are that before feedback (time one), peer and self rater perceptions align themselves more closely than after the initial feedback. Time two assessments show more alignment between subordinate and self rater perceptions. This finding is an overall assessment. It implies that there are differences in perceptions between peer and subordinate ratings at time one and time two. Additionally, there are differences in agreement between peer and subordinate ratings at time one and time two as measured by direct comparison of mean scores, which is indirectly supported by non-statistically significant ANCOVA results. This clearly implies differences in the way peers and subordinates rate leaders in the FBI work setting, and lends further credibility to all the findings in this study regarding peer and subordinate distinctions.

Further, these findings not only suggest that in the FBI culture peers perceive their leaders differently than do subordinates but also that they perceive them as better leaders. As discussed, these findings may have to do with the difference vantage points and knowledge that each of these groups has about the challenges leaders face. The context in which each of these groups finds itself with the accompanying responsibilities may be the significant factor that causes each to rate leaders differently.

### Implications

There are many implications for the present study. First, it is evident that perceptions are highly subjective, malleable, and unique to each rater involved. Clearly, individuals' perceptions about others are dependent upon the context and position from which they must observe. Using multi-rater feedback as part of an overall performance development system embraces this understanding. In fact, Day (2000) considers this one of the positive aspects of the multi-rater feedback process. By receiving feedback from more than one layer within an organization with whom the individual being rated regularly interacts, the supervisor's overall understanding of how he/she affects others should be enhanced. While this study did not clearly evidence that leaders in the FBI do



change their overall understanding of how they affect others as an overall scale (Research Question 1), there was non-statistically significant evidence through the sub-hypotheses that in various leadership dimension sub-scales of the LCCIS time two difference in means were in greater alignment than time one difference of means. This implies that in some instances the rated individuals may have improved their understanding of how they affect others as reflected in their ratings of themselves at time two.

Further implications for this research refutes the work of Jones and Nisbett (1972) who found that differences between self-perception and others' perceptions do exist. Their findings may suggest individuals attribute their behavior to situational circumstances, while observers tend to see others' behaviors in terms of internal dispositional explanations. A common assumption is that individuals have greater episodic knowledge of the self than do others (Locke, 2002). Supervisors in this study may also have seen their reasons for doing things as situational, and as separate groups, subordinates and peers may have seen them as permanent aspects of their personality and character, including their leadership credibility. When combined into "others" however, the results of ratings from subordinates and peers showed no statistical differences between self-perception and others' perceptions.

Here it was hypothesized that individuals would rate themselves more highly than their peers and subordinates on their initial multi-rater assessments. The findings suggest that the individuals being rated generally perceive themselves differently from both their peers and their subordinates when these two groups are bifurcated (Research Question 3). Generally, self-raters rated themselves lower than their peers and higher than their subordinates. As an overall assessment (Research Question 1) however, self-raters and others (peers and subordinates together) rated themselves similarly. The implications are that subordinates and peers see the rated individuals' reasons for doing things differently (Question 6) and the rated individuals have their own perspective. This may be due to situational contexts based on their position in the FBI work environment or due to the experiential understanding of peers versus subordinates.

With regard to the two subscales where there were perhaps the most interesting findings, MANAGE DIVERSITY and BUILD PUBLIC TRUST, in all instances the results showed exceedingly high ratings by peers, subordinates, and self raters (as

measured by an average mean score). For example the highest mean average at time one for self was 4.7412 in the category BUILD PUBLIC TRUST. The second highest was 4.7260 in the category MANAGE DIVERSITY. These numbers imply very high self-ratings in each of these categories and reflect a positive self-image by those in this study in these categories. The time two self ratings are lower but are still higher than all of the other time two self ratings in this study at 4.6743 and 4.6849, respectfully.

When others' (peers and subordinates together) ratings are considered across these categories the findings remain consistent. Others did not rate the leaders as highly as they rated themselves in the categories BUILD PUBLIC TRUST and MANAGE DIVERSITY, but the scores at time one were the first and second highest of all dimensions. The findings remain consistent at time two. When the mean scores of peers are taken into consideration the results are similar (BUILD PUBLIC TRUST and MANAGE DIVERSITY) at time one and time two, respectively. The only higher score was a time one mean score in the category CHALLENGE and SUPPORT the ORGANIZATION. The results also were similar with regard to subordinates' ratings at time one and time two. The highest ratings were in the category BUILD PUBLIC TRUST, the second highest were in the category CHALLENGE and SUPPORT the ORGANIZATION, and the third highest were in the category MANAGE DIVERSITY. Clearly, the categories BUILD PUBLIC TRUST and MANAGE DIVERSITY are significant for the FBI work setting. They reflect a special relationship with the American public and the unique responsibilities of a law enforcement service. These two categories of the instrument were developed by the contractor in conjunction with the FBI as part of the LCCIS in an attempt to tap these attributes since they are thought to be critical for success in law enforcement. It is both pleasing and reassuring to see that leaders in the FBI rate themselves and others very highly in these important leadership dimensions.

Finally, while Hazucha et al. (1993) found that there is evidence that managerial performance may improve following multi-rater feedback, in this case there was no statistically significant change (as measured by a Z score, t tests and through ANCOVA testing) between first and second assessments as an overall measurement (Hypothesis 1). Further, in time one versus time two difference of means comparisons for all hypotheses,

t tests showed no statistically significant differences. This implies that within the FBI work setting there is no evidence that multi-rater feedback has a significant effect on improving performance as measured by the perception of peers and subordinates within the system that currently exists and within the limitations of this study. However, in certain dimensions there is some non-statistically significant evidence of more alignment at time two assessments than time one assessments.

Clearly, this study showed that leadership in the FBI is strong as measured by mean averages of self and others on a five point Likert scale. Additionally, it showed that in two unique dimensions of the LCCIS, MANAGE DIVERSITY and BUILD PUBLIC TRUST, there were exceptionally high results. These findings are reassuring given the importance of the work the FBI does.

With regard to the use of the LCCIS within the organization, this study showed evidence that as but one piece of the leadership development process there are some limitations with it. While very useful in providing feedback about the perceptions of others that may have had a profound effect on a very small number of participants, the instrument by itself can only offer a partial understanding to any individual about how they can improve in their leadership abilities.<sup>26</sup> Other training and learning opportunities are necessary for any person to maximize their skills in the context of adult learning and human resource development. In this case the LCCIS and its attendant idiosyncrasies must be considered as but one piece of this process.

### Limitations

As with all social science research, there are limitations to the present study. There are also factors that potentially caused findings from this study to be affected, shifted, or conclusions from results to be less generalizable. This section will delineate the more apparent limitations and biases.

One threat to the credibility of the present study is related to the distribution of the survey instrument. As indicated in Appendix D, the instruments were sent to the assessed individual with instructions to have one of their subordinates distribute them. Because of this process, there was the assumption that raters were anonymous and more objective

---

<sup>26</sup> The results of any one or two participants are unknown within the context of the overall results.

than those who the assessed individual might have chosen, and the potential for the halo effect was potentially minimized. Since this assumption was not independently tested, there is no way of knowing what effect distributing the materials this way and the choice of peer(s) and subordinate(s) selected may have had on the validity of responses.

Another potentially biasing factor was the difference in the number of raters who actually filled out the questionnaire for each assessed individual. Each individual was sent a package with 14 peer/direct report evaluations and one self-evaluation. It is assumed that all 15 of these instruments would be properly filled out and returned to the data collection center for collation. Obviously, results were skewed or atypical if too few of the instruments were returned on a large number of the participants. This research used 73 participants and most had four or more responses from both peers and subordinates. Therefore, there should have been enough data sets to minimize idiosyncratic responses. However, a broader-based study consistently using more participants per rater may have more statistical power and validity.

Because this process relied on peers and subordinates for most of the feedback, another potentially biasing area may be the distribution of instruments to peers versus subordinates. Varying percentages of peers versus subordinates may have a tendency to skew the response results. This potential bias was also not independently tested, and there is no way of knowing what effect this may have had on the outcome of those hypotheses that combined responses.

Another factor was that the subordinate and peer evaluation were done by different people for the first versus second assessment. As a result, some of the change indicated by the second assessment may be due to perceptions of different individuals.

Another consideration that could have been limiting is that because of the Likert scale response used in the selected instrument (1-5 rating), responses were limited. Therefore, a certain amount of depth was lacking in the feedback than would be the case with a broader scale.

A further limitation is that the pool of individuals used in this research was thoroughly screened prior to their entry into the FBI. As such, their tolerance for issues such as diversity and regard for public trust are possibly different than they would be in an unscreened civilian population, and the results obtained from this area of the research

must be scrutinized with this in mind. This screening of employees is inherent in most of law enforcement and may be a reason for limited research in this occupational field.

Finally, while the results may be universally interpreted, it is noted that the organizational context is likely to impinge on the leadership dimensions being noted by the instrument being used. The findings must be interpreted within the context of the study.

#### Recommendations for Further Research

Multi-rater feedback has been widely publicized. Yet, as discovered in the literature, it is perhaps one of the least understood management developments in recent years (Church & Bracken, 1997). Furthermore, it has been reported that little definitive research concerning this process exists (Jackson & Greller, 1998; Nowak et al., 1999). This research offered some insights into this human resource development practice. However, there are many other areas yet to explore within this concept.

Goal setting and its relationship to multi-rater feedback have not been sufficiently explored according to Brutus et al. (1998). These authors suggest that some certainty exists with regard to 360-degree feedback, and that is that multi-rater feedback affords the feedback recipient a range of information that can be used for goal setting. However, little is known about how managers receiving multi-rater feedback set goals to change their behavior. Clearly, once growth areas are identified as a result of feedback from subordinates and peers, there is an opportunity to learn about oneself and change behavior. For many, goal setting is a way of changing behavior, and is a method frequently used for improving performance.

Whether or not individual development plans that include a goal setting component are created or an individual sets goals on their own as a result of learning from feedback was an area not addressed here. Goal setting as a way to improve performance becomes apparent as a natural extension to this research based on the overall finding of no perceptual changes of leadership over time. The present research could be built on by having those rated set goals for themselves to improve in the various leadership dimensions of the LCCIS. The present system used by the FBI and similar systems utilized by other organizations could be augmented with personal developmental

plans, including a goal setting component. This concept will be addressed in the “Recommendations for Practice” section below.

First impression or primacy effects had implications for the present research, but were not addressed. The assumption was made that with 18 months of observation primacy effects would have worn away and would have little influence on the assessments made by those receiving the feedback. Additionally, by the very nature of the multi-rater process whereby many perspectives (subordinate and peer) were garnered, such biases would become minimal or non-existent. This concept was not illuminated in the present study due to the amount of time that each of the raters was able to observe those being rated. In this analysis each rater had the opportunity for at least a year and half to make observations about the people they rated. Further research may determine the length of time it takes for first impressions to “wear away.”

To add further depth to the concepts of episodic and situational assessments of self, Markus and Wurf (1987) suggest that when an individual is thinking about and describing him or herself, they do so with knowledge they have in their short-term memory. What they use to describe themselves is with the attributes they have in their working self-memory, which holds only a very restricted amount of information. This may also be true when individuals assess others. These aspects of the person perception process may have had implications for the present research, although they were not measured. With any part of the perception process there is a certain amount of subjectivity that is in the eye of the beholder. Even though those doing the rating in this instance had the opportunity to observe for an extended period, there is a possibility that recency effects or other perception attributes may have had an effect on their responses. The results suggested a fairly stable measure of those rated as perceived by others (Hypothesis 1) at both times. It is likely that episodic or situational assessments had a minimal effect on the present findings. However, these effects were not examined in this instance and offer another area suitable for study.

Fleenor, McCauley, and Brutus (1996) suggest that whether or not individuals rate themselves higher, lower or the same as others has no bearing on their job performance effectiveness. This is somewhat the opposite side of the present research, which addressed how raters perceived themselves or others in their work capacity. This

view of the person perception process would be interesting to explore in an effort to examine self-concept and its relevance to job performance. This idea was not being measured in the present research and is another area available for study.

One more area for research in the multi-rater feedback domain is the comparison of the quality of leadership by those who have had the benefit of more than one multi-rater feedback intervention. While it was not the focus of this research, an area for further investigation may be to compare leadership levels of those who have had the benefit of two versus one assessment. The results from this examination (as an overall measure) would support hypotheses that suggest no changes if these types of comparisons are made. Further study may indicate other findings.

This research was primarily quantitative in nature. Yet there is a place for qualitative research that focuses on multi-rater feedback. The multi-rater feedback instrument used in this study had a place for written comments, which were part of the assessment given to each feedback recipient. This study intentionally did not use any of this information both because its focus did not lend itself to this type of analysis and to ensure the anonymity of each of those receiving the feedback. Undoubtedly, there was rich information contained in these comments that would be valuable to analyze for research purposes. Further, this type of analysis could incorporate interviews allowing for greater understanding of those who gained most from the multi-rater feedback process.

The 360-degree feedback procedure consists of three components: data, evaluation, and action (Jackson & Greller, 1998). The data are the factual information obtained from the raters who observe the person being measured. These data usually take the form of responses to questions, or narrative descriptions, which raters provide on survey questionnaires. The evaluation element consists of a value judgment or reading of responses. The action element relates to what occurs as a result of the process. According to these two researchers, the action element may be the key to effective utilization of a 360-degree feedback system. This action element is another area that is open for further study. A subcomponent of this idea may be the goal setting concept mentioned earlier and could be incorporated into a personal developmental plan used in conjunction with the multi-rater feedback process.

Finally, Hooijberg and Choi (2000) point out that little focus has been given to the essence of the differentiation between self and other raters. This study's literature search did examine how self and others do perceive differently, but it is clear that the research is limited in this area. Additionally, while this research discovered that self and other raters (peers and subordinates together) perceive leaders basically the same, clearly as separate groups subordinates and peers perceive leaders differently. This discovery, which was not the central focus of this research, is an area that can be examined more extensively both conceptually and in practice. Further, why peers rated leaders higher than subordinates in this particular research is speculated about and is likely associated with the context from which they must observe. More research could help illuminate these particular findings and suppositions.

#### Recommendations for Practice

This research produced a variety of interesting findings. The implications are that leadership in the FBI, as measured by different dimensions, is strong. Yet, there is no evidence that as an overall measure leaders or others in this organization change their perceptions over time as measured by the LCCIS. It appears that there is no verification by the means used in this study that the current use of the multi-rater feedback instrument alone is having the effect of significantly improving the perception of leadership in the FBI. Consequently, there is the possibility that those attending the EDI I training are not doing the kinds of critical thinking about their leadership practice that Mezirow (1990) and Brookfield (1990) refer to when they discuss this concept, particularly as it relates to changed behavior following the reflective process. These findings may point to the need for multi-rater instruments to be used in concert with personal development plans where there is an expectation of personal reflection and changed behavior.

The multi-rater instrument used in this research was not tied to any kind of a developmental plan for the individual receiving the feedback. It is suggested that by attaching the instrument's use to this kind of individual strategy for personal growth, the latter results (time two) may be different. It would appear that this kind of a performance development system would be of benefit in this instance as in spite of the positive aspects of the use of the multi-rater assessment, possibly more could have been gained if there



was a developmental plan associated with it. This plan could include measurable goals either recorded in some formal or informal way by the organization or by each assessed individual. Further, the other leadership training and practices in use by the organization could be tied to the personal developmental plan as well. In this way there would be a comprehensive assessment that would be associated with a performance improvement strategy. This plan could be used as a basis for developmental opportunities garnered inside or outside the organization depending on the specific needs of the assessed.

A personal developmental plan has the commensurate associated costs that accompany it, and requires other ancillary considerations. First, it necessitates those administering the written feedback be trained as professionals in providing such guidance. To be credible, these trainers need the appropriate education and background that allows for those receiving the feedback to “buy in.” Also, these types of reports take considerable time to create and draft, and any organization interested in such services must be prepared to pay a fairly substantial sum for them. Next, it is critical that the culture of the organization embrace this kind of a feedback system and that it be supported by the leadership at the highest levels. As was noted earlier with regard to the adoption of the use of the multi-rater feedback instrument, once one of the senior leaders in the organization was evaluated with it, others bought in as well. Such support is critical for these systems to be accepted as cultural norms. Further, in order to obtain the maximum amount of integrity with this type of a system, the individuals receiving the feedback have to believe the information provided in written reports is confidential. Otherwise, there is possibility of challenge. This is especially true given the litigious nature that currently exists within the public and private sectors. Finally, an education process needs to take place whereby those participating in the multi-rater feedback process gain a full knowledge of what is expected of them before, during, and after receiving such feedback and their individual developmental plans.

It would be interesting to pursue research regarding the gains in the perceptions of leadership in organizations using individual development plans in conjunction with multi-rater feedback. After a close examination of the use of multi-rater feedback, particularly in this instance, it is reasonable to assume that if this kind of plan was in effect the results of this research may have been different.

## Conclusion

This chapter delineated a summary of the study presented here. It began by providing a summary overview, followed by a discussion of the study's conclusions, implications, and limitations. Finally, this chapter concluded with recommendations for further research and practice. Through this sequence, an intricate look into a public sector law enforcement organization, the FBI, was discovered. Further, a rich and highly developed understanding of the leadership as evaluated from those within the organization was uncovered. The results shed further light on organizational systems instruments used by human resource professionals in the form of the multi-rater feedback process.

## References

- Asch, S. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 48*, 258-290.
- Ashford, S., & Tsui, A. (1991). Self-regulation for managerial effectiveness: The role of active feedback seeking. *Academy of Management Journal, 34*, 251-280.
- 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. E., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology, 48*, 35-59.
- Atwater, L., & Waldman, D. (1998, May). Accountability in 360-degree feedback. *HR Magazine, 43*, 96-98.
- Atwater, L. E., Waldman, D.A., Atwater, D., & Cartier, P. (2000). An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology, 53*, 275-297.
- Atwater, L.E., & Yammarino, F.J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141-164.
- Atwater, L.E., & Yammarino, F.J. (1997). Self – other rating agreement: A review and model. *Research in Personnel and Human Resource Management, 15*, 121-174.
- Bennis, W., & Nanus, B. (1985). *Leaders: the strategies for taking charge*. New York, NY: Harper & Row.
- Bennis, W., & Townsend, R. (1995). *Reinventing leadership: Strategies to empower the organization*. New York, NY: William Morrow.
- Bernardin, J. H., & Beatty, R.W. (1987). Can subordinate appraisals enhance managerial productivity? *Sloan Management Review, 28*, 63-73.
- Bernardin, J.H., Dahmus, S. A., & Redmon, G. (1993). Attitudes of first-line supervisors toward subordinate appraisals. *Human Resource Management, 32*, 315-324.
- Bernstein, B.L., & Lecomte, C. (1979). Supervisory-type feedback effects: Feedback discrepancy level, trainee psychological differentiation, and immediate responses. *Journal of Counseling Psychology, 26*, 295-303.

- Bohl, D.L. (1996, September). Minisurvey: 360-degree appraisals yield superior results, survey shows. *Compensation & Benefits Review*, 28, 16-19.
- Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M. Dunnette & L. Hough (Eds.). *Handbook of industrial and organizational psychology (2<sup>nd</sup> ed.)*, Vol. 2. Palo Alto, CA: Consulting Psychology Press.
- Bracken, D. W., Dalton, M.A., Jako, R., McCauley, C.D., & Pollman, V.A. (1997). *Should 360-degree feedback be used only for development purposes?* Greensboro, NC: Center for Creative Leadership.
- Brown, J.D. (1986). Evaluation of self and others: Self-enhancement biases in social judgments. *Social Cognition*, 4, 353-376.
- Brookfield, S. (1990). Using critical incidents to explore learners' assumptions. In Mezirow, J., & Associates (Eds.) *Fostering critical reflection in adulthood* (pp. 177-193). San Francisco, CA: Jossey-Bass.
- Brutus, S., Fleenor, J., & McCauley, C. D. (1996, April). *Self-other rating discrepancy in 360-degree feedback: An investigation of demographic and personality predictors*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, San Diego.
- Brutus, S., Fleenor, J. & London, M. (1998). Elements of effective 360-degree feedback. In W. Tornow, M. London, & CCL Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 11-27). San Francisco, CA.: Jossey-Bass.
- Budman, M., & Rice, B. (1994, February). The rating game: Performance appraisal. *Across the Board*, 31, 34.
- Church, A. H., & Bracken, D.W. (1997, June). Advancing the state of the art of 360-degree feedback: guest editor's comments on the research and practice of multi-rater assessment methods. *Group and Organization Management*, 22, 149-161.
- Clavell, J. (1983). *The art of war by Sun Tzu*. New York, NY: Delacorte.
- Cohen, W.A. (1990). *The art of the leader*. Paramus, NJ: Prentice Hall.
- Cohen, W. A. (2002). *The new art of the leader - Leading with integrity and honor*. Paramus, NJ: Prentice Hall.
- Covey, S. R. (1989). *The seven habits of highly effective people*. New York, NY: Simon & Schuster.

- Covey, S. R. (1991). *Principle centered leadership*. New York, NY: Simon & Schuster.
- Dalton, M. A. (1998). Five rationales for using 360-degree feedback in organizations. In W. Tornow, M. London, & CCL Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 59-77). San Francisco, CA: Jossey Bass.
- Dalton, M. & Hollenbeck, G. P. (1996). *How to design an effective system for developing managers and executives*. Greensboro, NC: Center for Creative Leadership.
- Day, D.V. (2000). Leadership development: A review in context. *Management Quarterly*, 11, 581-613.
- Delattre, E. (1994). *Character & cops*. Washington, DC: The American Enterprise Institute.
- DeLoach, C. (1995). *Hoover's FBI*. Washington, DC: Regnery.
- De Vries, M.F.R. (1994). The leadership mystique. *Academy of Management Executive*, 8, 73-92.
- Donnithorne, L. R. (1993). *The West Point way of leadership*. New York, NY: Doubleday.
- Dunnette, M.D. (1993). My hammer or your hammer? *Human Resource Management*, 32, 373-384.
- Drucker, P.F. (1996). Foreword. In F. Hesselbein, M. Goldsmith, & R. Beckhard (Eds.), *The leader of the future* (pp. xi-xv) San Francisco, CA: Jossey-Bass.
- Edwards, M.R., & Ewen, A.J. (1996). *360-degree feedback*. New York, NY: American Management Association.
- Federal Bureau of Investigation. (n.d.). *Mission Statement of the FBI*. Retrieved May 21, 2004, from <http://kansascity.fbi.gov/mission.htm>.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Fleenor, J., McCauley, C., & Brutus, S. (1996). Self – other rating and leader effectiveness. *Leadership Quarterly*, 7, 487-506.
- Freemont, California Police Department 360-degree Feedback Form. (n.d.).
- Freire, P. (1970). *Pedagogy of the oppressed*. New York, NY: Seabury Press.

- Goldsmith, M. (1996, Fall). Coaching for behavioral change. In *Leader to Leader*, 2. No Publisher cited.
- Goldsmith, M. (1997, Summer). On a consumer watershed. In *Leader to Leader*, 5. No Publisher cited.
- Halman, F., & Fletcher, C. (2000). The impact of development center participation and the role of individual differences in changing self-assessments. *Journal of Occupational and Organizational Psychology*, 73, 423-442.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self – supervisor, self – peer and peer – supervisor ratings. *Personnel Psychology*, 41, 43-62.
- Hazucha, J., Hezlett, S., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills and development. *Human Resource Management*, 32, 325–351.
- Hooijberg, R., & Choi, J. (2000). Which leadership roles matter to whom? An examination of rater effects on perceptions of effectiveness. *The Leadership Quarterly*, 11, 341-364.
- Jackson, J. H., & Greller, M. M. (1998). Decision elements for using 360 (degree) feedback. *Human Resource Planning*, 21, 18-22.
- Jones, E.E., & Nisbett, R.E. (1972). The actor and observer: Divergent perceptions of the causes of behavior (pp. 79-94). In E.E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.
- Kaltman, A. (2000). *The leadership lessons of Robert E. Lee*. Paramus, NJ: Prentice Hall.
- Keilty, Goldsmith, & Co. (2001). Letter from John F. Greene, Ph.D. to FBI Academy, Leadership and Management Science Unit.
- Kirk, R.E. (1968). *Experimental design: Procedures for the Behavioral Sciences*. Belmont, CA: Brooks/Cole Publishing.
- Kluger, A.N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback theory. *Psychological Bulletin*, 119, 254-284.
- Knapp, M.L. (1978). *Social intercourse*. Boston, MA: Allyn and Bacon, Inc.
- Kotter, J.P., & Heskett, J.L. (1992). *Corporate culture and performance*. New York, NY: Free Press.

- Kouzes, J., & Posner, B. (1987). *The leadership challenge: How to get extraordinary things done in organizations*. San Francisco, CA: Jossey-Bass.
- Kouzes, J., & Posner, B. (1993). *Credibility*. San Francisco, Ca.: Jossey-Bass.
- Kouzes, J., & Posner, B. (1995). *The leadership challenge: How to keep getting extraordinary things done in organizations*. San Francisco, CA.: Jossey-Bass.
- Kreitner, R., & Kinicki, A. (2003). *Organizational behavior (6<sup>th</sup> ed.)*. New York, NY: McGraw Hill.
- Landy, F.L., & Farr, J.L. (1983). *The measurement of work performance*. San Diego, CA: Academic Press.
- Laborde, G.Z. (1997). *Influencing with integrity*. Palo Alto, CA: Syntony Publishing.
- Locke, K.D. (2002). Are descriptions of the self more complex than descriptions of others? *Personality and Social Psychology Bulletin*, 28, 1094-1105.
- Loehr, J. (1997). *Stress for success*. New York, NY: Random House.
- London, M. & Beatty, R.W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management*, 32, 353-372.
- London, M. (1997). *Job feedback: Giving, seeking, and using feedback for performance improvement*. Mahwah, NJ: Earlbaum.
- London, M & Smither J. W. (1995). Can multi-source feedback change self-evaluations, skill development, and performance? Theory-based applications and directions for research. *Personnel Psychology*, 48, 803-839.
- London, M., & Tornow, W.W. (1998). 360-degree feedback – more than a tool. In W. Tornow, M. London, & CCL Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 1-8). San Francisco, CA: Jossey Bass.
- London, M., & Wohlers, A.J. (1991). Agreement between subordinate and self-ratings in upward feedback. *Personnel Psychology*, 44, 375-390.
- Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, 38, 299-337.
- Martineau, J. (1998). Using 360-degree surveys to assess change. In Tornow W., London, M. & CCL Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 217-248). San Francisco, CA: Jossey-Bass.
- Maxwell, J.C. (1993). *Developing the leader within you*. Nashville, TN: Thomas Nelson.

- Maxwell, J.C. (1998). *The 21 irrefutable laws of leadership: Follow them and people will follow you*. Nashville, TN: Thomas Nelson.
- Meyer, H. (1980). Self-appraisal of job performance. *Personnel Psychology*, 33, 291-295.
- Mezirow, J. (1990). Conclusion: Toward transformative learning and emancipatory education. In J. Mezirow & Associates (Eds.), *Fostering critical reflection in adulthood: A guide to transformative and emancipatory education* (pp. 354-376). San Francisco, CA: Jossey-Bass.
- Murphy, K.R., & Cleveland, J.N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Nadler, D. A. (1977). *Feedback and organization development: Using data based methods*. Reading, MA: Addison-Wesley.
- Nadler, D. & Tushman, M. (1990). *Qualitative data analysis: An expanded sourcebook (2<sup>nd</sup> ed.)*. Thousand Oaks, CA: Sage.
- Nelson, T.D. (1993). The hierarchical organization of behavior: A useful feedback model of self-regulation. *Current Directions in Psychological Science*, 2, 121-126.
- Nowak, K.M., Hartley, J., & Bradley, W. (1999, April). How to evaluate your 360 feedback efforts. *Training and Development*, 53, 48-53.
- Oakley, & Krug, D. (1991). *Enlightened leadership*. New York, NY: Simon & Schuster.
- Oceanside, California Police Department Subordinate Feedback Form. (n.d.).
- Pearce, J.L. & Porter, L.W. (1986). Employee responses to formal performance appraisal feedback. *Journal of Applied Psychology*, 71, 211-218.
- Pedler, M. & Boydell, T. (1980). Is all management development self-development? In J. Beck & C. Cox (Eds.), *Advances in management education* (pp. 165-196). New York, NY: Wiley.
- Reilly, R.R., Smither, J.W., & Vasilopoulos, N.L. (1996). A longitudinal study of upward feedback. *Personnel Psychology*, 49, 599-612.
- Peters, T., & Waterman, R., Jr. (1982). *In search of excellence*. New York, NY: Warner Books.
- Schein, E. (1992). *Organizational culture and leadership*. San Francisco, Ca.: Jossey-Bass.



- Schneider, D.J., Hastorf, A. H., & Ellsworth, P.C. (1979). *Person perception*. Reading, MA: Addison-Wesley.
- Shore, L.M., & Thornton, G.C. (1986). Effects of gender on self- and supervisory ratings. *Academy of Management Journal*, 29, 115-129.
- Smither, J. W., Wohlers, A.J., & London, M. (1995). A field study of reactions to normative versus individualized upward feedback. *Group and organization management*, 20, 61-89.
- Spencer, L.M., & Spencer, S.M. (1993). *Competence at work: models for superior performance*. New York, NY: Wiley.
- Staub, R. E. (1996). *The heart of leadership: 12 practices of courageous leaders*. Provo, UT: Executive Excellence.
- Sternbergh, W.W. (1998). 360 in Action: Following Catharine's footsteps through a leadership development program. In W. Tornow, M. London, & CCL Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 28-56). San Francisco, CA.: Jossey Bass.
- Stubblefield, H. & Keane, P. (1994). *Adult education in the American experience*. San Francisco, CA: Jossey-Bass.
- Tabachnick, B.G. & Fidell, L.S. (1996). *Using multivariate statistics (3<sup>rd</sup> ed.)*. New York: Harper Collins.
- Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology*, 33, 263-271.
- Tichy, N. M. (1997). *The leadership engine*. Harper Business
- Tornow, W. (Ed.) (1993). Editor's note: Introduction to special issue on 360-degree feedback. Special issue on 360-degree feedback. (pp. 211-219). *Human Resource Management*, 32, 2-3.
- Tornow, W., London, M., & CCL Associates (1998). *Maximizing the value of 360-degree feedback*. San Francisco, CA: Jossey-Bass.

- Van Velsor, E. (1998). Designing 360-degree feedback to enhance involvement, self-determination, and commitment. In W. Tornow, M. London & CCL Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 149-195). San Francisco, CA: Jossey-Bass.
- Van Velsor, E. & Leslie, J. B. (1991). Feedback to managers, *Volume II: A review and comparison of sixteen multi-rater feedback instruments*. Greensboro, N.C.: Center for Creative Leadership.
- Walker, H.M., & Lev, J. (1969). *Elementary statistical methods (3<sup>rd</sup> ed.)*. New York, NY: Holt, Rinehart, and Winston, Inc.
- Wanguri, D. M. (1995). A review, an integration, and a critique of cross-disciplinary research on performance appraisals, evaluations, and feedback: 1980-1990. *Journal of Occupational and Organizational Psychology*, 66, 3-11.
- Weber, G. B. (1996). Growing tomorrow's leaders. In F. Hesselbein, M. Goldsmith, & R. Beckhard (Eds.), *The leader of the future* (pp. 303-309). San Francisco, CA: Jossey-Bass.
- Wohlers, A.J., Hall, M.J., & London, M. (1993). Subordinates rating managers: Organizational and biographical correlates of self/subordinate agreement. *Journal of Occupational and Organizational Behavior*, 66, 263-275.
- Warech, M.A., Smither, J.W., Reilly, R.R., Millsap, R.E. and Reilly, S.P., 1998. Self-monitoring and 360-degree ratings. *Leadership Quarterly*, 9, 449-473.
- Wycoff, M. A. & Oettmeier, T.N. 1994. *Evaluating patrol officer performance under community policing: The Houston experience*. Washington, DC: National Institute of Justice.
- Yammarino, F., & Atwater, L. (1993). Understanding self-perception accuracy: Implications for human resources management. *Human Resource Management*, 32, 231-247.
- Yukl, G. (2002). *Leadership in Organizations (5<sup>th</sup> ed.)*. Upper Saddle River, NJ: Prentice Hall.

Yukl, G., & Van Fleet, D. D. (1992). Theory and research on leadership in organizations.  
In M. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational  
psychology (2<sup>nd</sup> ed.)*. Palo Alto, CA: Consulting Psychologists.

# APPENDIX A

# LEADERSHIP COMMITMENTS AND CREDIBILITY INVENTORY SYSTEM



**SELF  
FORM**

**Print Your Name**

## GENERAL INSTRUCTIONS

This inventory system is designed to measure the effectiveness of your leadership practices as rated by you and others (peers, direct reports). Peers are defined as people organizationally on your level. Direct reports are people you supervise. The goal is to have seven of each (6/8 or 5/9 is permissible). If you have a current assignment without sufficient direct reports, then you may distribute instruments to those who worked with you in the recent past.

First, print your name on the cover of all forms. On the *Other Form*, check the peer box on forms you plan to give to peers and check the direct report box on forms you plan to give to direct reports. Then, ask your peers and/or direct reports to complete the *Other Form* for you. Be certain to assure them their ratings are anonymous; encourage them to be candid. To assure validity and anonymity, a minimum of three reports must be received from Peers and three from Direct Reports.

Next, complete this self form. Record your perceptions of how others see you. To help understand each item, read it in context with the group heading.

You are the only person in the FBI who will see the results of this inventory system in a comprehensive summary. Norms for 200 or more anonymous people will be provided the FBI Leadership and Management Science Unit for training development needs analysis. The feedback you will receive is for your personal self-development.

Be certain all forms are completed **within 3 days** and mailed to:

**NCS Organizational Assessment Systems  
4401 West 76th St.  
Edina, MN 55435.**

## LEADERSHIP COMMITMENTS AND CREDIBILITY

Instructions: As you complete this questionnaire, please note that each item is preceded with the question, “How satisfied are others with your ability to . . .” Your response choices are described at the top of each page. Please indicate your choice by **circling** your selection for each numbered item.

..... ANSWER OPTIONS — CIRCLE ONE					
<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

### STEPS TO EMPOWERMENT

**Provide Enabling Structure:** Note: Structure is coaching and teaching how to do a task.

1. Provide enabling structure when needed ..... **HD D N S HS UR**
2. Demonstrate how to do specific tasks when necessary ..... **HD D N S HS UR**
3. Make certain that goals are understood ..... **HD D N S HS UR**
4. Change from a “boss” to a “coach” ..... **HD D N S HS UR**
5. Provide appropriate orientation on new assignments ..... **HD D N S HS UR**
6. Provide timely coaching and feedback ..... **HD D N S HS UR**

### Support and Collaborate

7. Collaborate with others in setting goals ..... **HD D N S HS UR**
8. Encourage others to participate in making decision ..... **HD D N S HS UR**
9. Encourage others to discuss their concerns ..... **HD D N S HS UR**
10. Support others when they need it ..... **HD D N S HS UR**
11. Motivate others by building their confidence ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**Empower Others**

- 12. Eliminate unnecessary sign-offs (approvals)..... **HD D N S HS UR**
- 13. Delegate authority as well as responsibility ..... **HD D N S HS UR**
- 14. Empower others to make decisions ..... **HD D N S HS UR**

**Analyze Others' Readiness** Note: readiness is ability, willingness, and confidence to do a task.

- 15. Break the work into specific tasks ..... **HD D N S HS UR**
- 16. Assess the ability of others to do specific tasks ..... **HD D N S HS UR**
- 17. Assess the motivation of others to do specific tasks ..... **HD D N S HS UR**
- 18. Ask others what they need to do their job better ..... **HD D N S HS UR**
- 19. Avoid over-leading (too much structure) ..... **HD D N S HS UR**
- 20. Avoid under-leading (too little structure) ..... **HD D N S HS UR**

**CREDIBILITY**

**Competence**

- 21. Perform the technical/functional skills to do the job ..... **HD D N S HS UR**
- 22. Manage people ..... **HD D N S HS UR**
- 23. Manage conflict effectively ..... **HD D N S HS UR**



ANSWER OPTIONS — CIRCLE ONE

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .

**Character**

- 24. Ensure honest reporting of results ..... **HD D N S HS UR**
- 25. Follow through on commitments ..... **HD D N S HS UR**
- 26. Communicate honestly with others (No “hidden agendas”) ..... **HD D N S HS UR**
- 27. Maintain high integrity in all situations ..... **HD D N S HS UR**
- 28. Avoid “playing favorites” ..... **HD D N S HS UR**

**Composure**

- 29. Express confidence in difficult situations ..... **HD D N S HS UR**
- 30. Maintain composure under pressure ..... **HD D N S HS UR**
- 31. Encourage and accept candid criticism ..... **HD D N S HS UR**
- 32. Demonstrate a “can do” attitude (even if others say “it can’t be done”) ..... **HD D N S HS UR**
- 33. Improve as a result of receiving feedback from this questionnaire ..... **HD D N S HS UR**

**Courage**

- 34. Admit mistakes readily ..... **HD D N S HS UR**
- 35. Practice what you preach ..... **HD D N S HS UR**
- 36. Demonstrate courage to stand up for beliefs ..... **HD D N S HS UR**
- 37. Challenge higher level management when appropriate ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**Care For People**

- 38. Realize how he or she affects others ..... **HD D N S HS UR**
- 39. Treat others with respect and dignity ..... **HD D N S HS UR**
- 40. Avoid unnecessary destructive comments about others ..... **HD D N S HS UR**
- 41. Support an appropriate balance between work and non-work life  
for others ..... **HD D N S HS UR**

**CHALLENGE and SUPPORT the ORGANIZATION**

- 42. Represent the organization in a positive manner ..... **HD D N S HS UR**
- 43. Support higher level management ..... **HD D N S HS UR**
- 44. Avoid unnecessary destructive comments about the organization  
and higher level management ..... **HD D N S HS UR**
- 45. Act proud to be a member of the organization ..... **HD D N S HS UR**
- 46. Appropriately challenge decisions you think are not good for  
the organization ..... **HD D N S HS UR**
- 47. Continuously strive to improve the organization ..... **HD D N S HS UR**

**BUILD TEAMS**

- 48. Actively promote teamwork in your area ..... **HD D N S HS UR**
- 49. Share information across organizational boundaries ..... **HD D N S HS UR**
- 50. Establish relationships across a wide range of areas  
throughout the organization ..... **HD D N S HS UR**
- 51. Promote other areas of the organization ..... **HD D N S HS UR**

ANSWER OPTIONS — CIRCLE ONE

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .

**DEVELOP INDIVIDUALS**

- 52. Provide developmental feedback on performance problems ..... **HD D N S HS UR**
- 53. Give positive feedback for improvement in performance ..... **HD D N S HS UR**
- 54. Provide growth opportunities for others ..... **HD D N S HS UR**
- 55. Support people of all backgrounds to reach their full potential  
(regardless of race, gender, or other differences) ..... **HD D N S HS UR**
- 56. Provide candid performance feedback in a timely manner ..... **HD D N S HS UR**
- 57. Treat employee mistake as opportunity for instruction ..... **HD D N S HS UR**
- 58. Correct employee's misconduct ..... **HD D N S HS UR**

**QUALITY and TASK**

- 59. Maintain high quality standards ..... **HD D N S HS UR**
- 60. Communicate the importance of quality ..... **HD D N S HS UR**
- 61. Strive for continuous improvement of relevant work processes ..... **HD D N S HS UR**
- 62. View change as an opportunity ..... **HD D N S HS UR**
- 63.. Accept and manage risk ..... **HD D N S HS UR**
- 64. Originate action to complete the task ..... **HD D N S HS UR**
- 65. Accomplish desired results with limited resources when necessary ..... **HD D N S HS UR**
- 66. Keep the right focus ..... **HD D N S HS UR**
- 67. Make the work meaningful and relevant ..... **HD D N S HS UR**
- 68. Explain how the work results in organizational success ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**CUSTOMER/CLIENT** Note: a customer or client is any person, internal or external, who receives a service or product. For instance, the following people could be an FBI employee’s customer: Fellow employee, crime victim, accused, witness, defense attorney, and prosecutor.

- 69. Follow through on commitments made to customers (see definition above) ..... **HD D N S HS UR**
- 70. Avoid sarcastic comments when customers complain ..... **HD D N S HS UR**
- 71. Listen to the ideas of customers for improving work product or services ..... **HD D N S HS UR**
- 72. Consider customers as a top priority ..... **HD D N S HS UR**
- 73. Deal with customer concerns on a timely basis ..... **HD D N S HS UR**
- 74. Assess customer satisfaction levels ..... **HD D N S HS UR**

**VISION**

- 75. Explain the organizational vision to others ..... **HD D N S HS UR**
- 76. Communicate a clear vision for the work group ..... **HD D N S HS UR**
- 77. Motivate others to commit to the vision ..... **HD D N S HS UR**
- 78. Establish a strategy to realize the vision ..... **HD D N S HS UR**
- 79. Implement actions in the work group that are in alignment with the organizational vision ..... **HD D N S HS UR**

**FOSTER CREATIVITY**

- 80. Encourage others to creatively solve problems ..... **HD D N S HS UR**
- 81. Avoid comments that stifle creativity (killer phrases) ..... **HD D N S HS UR**
- 82. Make reasonable risks to implement new ideas ..... **HD D N S HS UR**
- 83. Allow new ideas to be explored and evaluated ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**MANAGE DIVERSITY**

- 84. Provide equal opportunity to compete for assignments or promotions ..... **HD D N S HS UR**
- 85. Create a workplace environment free of intimidation ..... **HD D N S HS UR**
- 86. Create a workplace environment free of hostility ..... **HD D N S HS UR**
- 87. Create a workplace environment free of offensive behavior ..... **HD D N S HS UR**
- 88. Deal with complaints of discrimination effectively ..... **HD D N S HS UR**

**BUILD PUBLIC TRUST** Note: Government service implies unique responsibilities. Since the clients of Government employees are taxpayers, the services rendered are different from private industry. Taxpayer confidence in government employees is interpreted as “Public Trust.” Also, one who holds a position of “Public Trust” has a fiduciary or trustee type obligation to “The People.”

**Public Trust**

- 89. Use position only for legitimate personal gain ..... **HD D N S HS UR**
- 90. Ensure others use their position only for legitimate personal gain ..... **HD D N S HS UR**
- 91. Use taxpayers’ funds wisely ..... **HD D N S HS UR**
- 92. Be a team-player when working with other government employees ..... **HD D N S HS UR**
- 93. Assure the public has fair access to services ..... **HD D N S HS UR**

**Integrity**

- 94. Behave the same in private and public life ..... **HD D N S HS UR**
- 95. Exhibit moral behavior consistent with organizational values ..... **HD D N S HS UR**
- 96. Always uphold the Constitution ..... **HD D N S HS UR**
- 97. Always uphold the law ..... **HD D N S HS UR**

**Thank you**



Developed with Keilty, Goldsmith & Company  
for use in FBI management development programs

**IMPORTANT**

*Please mail completed form within three days to:*

*NCS Organization Assessment Systems  
4401 West 76th Street  
Edina, MN 55435*

MSU-WHS: 4.28.04

# APPENDIX B

# LEADERSHIP COMMITMENTS AND CREDIBILITY INVENTORY SYSTEM



**This form is being completed for:**

*Print Name of Individual being Evaluated*

## OTHER FORM

Are you this persons (Check ✓ one) **PEER**

Or **DIRECT REPORT**



## **GENERAL INSTRUCTIONS FOR PEERS AND DIRECT REPORTS**

**This inventory system is designed to measure and improve the leadership practices of the individual listed on this cover. Your candid responses are important. Please complete as many items as possible. To help understand each item, read it in context with the group heading. If you have insufficient information upon which to base a decision, circle Unable to Respond. Also, you are requested to provide written comments at the back of this booklet.**

**Do not identify yourself on this form. Your responses are anonymous. They will be combined with the responses of others to produce a comprehensive summary report for the person being assessed. The results are intended to show areas of opportunity for personal development.**

**The person you are rating will receive the summary confidentially. The only results provided the FBI will be norms for 200 or more anonymous people. The Leadership and Management Science Unit will use these norms to analyze training development needs.**

**Be certain that your relationship to the individual for whom you are completing this form is accurately indicated on the front cover (as Direct Report or Peer). Peers are people who are organizationally on the same level. Direct reports are people directly supervised, now or in recent past, by person being rated.**

**Please complete form within 3 days and mail to:**

**NCS Organization Assessment Systems  
4401 West 76th St.  
Edina, MN 55435**

**Thank you**

## LEADERSHIP COMMITMENTS AND CREDIBILITY

**Instructions:** As you complete this questionnaire, please note that each item is preceded with the question, “How satisfied are you with this persons ability to . . . .” Your response choices are described at the top of each page. Please indicate your choice by circling your selection for each numbered item.

ANSWER OPTIONS — CIRCLE ONE					
<i><b>HD</b></i>	<i><b>D</b></i>	<i><b>N</b></i>	<i><b>S</b></i>	<i><b>HS</b></i>	<i><b>UR</b></i>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

### STEPS TO EMPOWERMENT

**Provide Enabling Structure: Note: “Enabling Structure” includes coaching and teaching how to do a task.**

1. Provide enabling structure when needed ..... **HD D N S HS UR**
2. Demonstrate how to do specific tasks when necessary ..... **HD D N S HS UR**
3. Make certain that goals are understood ..... **HD D N S HS UR**
4. Change from a “boss” to a “coach” ..... **HD D N S HS UR**
5. Provide appropriate orientation on new assignments ..... **HD D N S HS UR**
6. Provide timely coaching and feedback ..... **HD D N S HS UR**

### Support and Collaborate

7. Collaborate with others in setting goals ..... **HD D N S HS UR**
8. Encourage others to participate in making decisions ..... **HD D N S HS UR**
9. Encourage others to discuss their concerns ..... **HD D N S HS UR**
10. Support others when they need it ..... **HD D N S HS UR**
11. Motivate others by building their confidence ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b> <i>Highly Dissatisfied</i>	<b>D</b> <i>Dissatisfied</i>	<b>N</b> <i>Neither Satisfied nor Dissatisfied</i>	<b>S</b> <i>Satisfied</i>	<b>HS</b> <i>Highly Satisfied</i>	<b>UR</b> <i>Unable to respond</i>
---	---------------------------------	---	------------------------------	--	---

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**Empower Others**

- 11. Eliminate unnecessary sign-offs (approvals) ..... **HD D N S HS UR**
- 12. Delegate authority as well as responsibility ..... **HD D N S HS UR**
- 13. Empower others to make decisions ..... **HD D N S HS UR**

**Analyze Others' Readiness Note: readiness is ability, willingness, and confidence to do a task.**

- 14. Break the work into specific tasks ..... **HD D N S HS UR**
- 15. Assess the ability of others to do specific tasks ..... **HD D N S HS UR**
- 16. Assess the motivation of others to do specific tasks ..... **HD D N S HS UR**
- 17. Ask others what they need to do their job better ..... **HD D N S HS UR**
- 18. Avoid over-leading (too much structure) ..... **HD D N S HS UR**
- 19. Avoid under-leading (too little structure) ..... **HD D N S HS UR**

**CREDIBILITY**

**Competence**

- 21. Perform the technical/functional skills to do the job ..... **HD D N S HS UR**
- 22. Manage people ..... **HD D N S HS UR**
- 23. Manage conflict effectively ..... **HD D N S HS UR**

ANSWER OPTIONS — CIRCLE ONE

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .

**Character**

- 24. Ensure honest reporting of results ..... **HD D N S HS UR**
- 25. Follow through on commitments ..... **HD D N S HS UR**
- 26. Communicate honestly with others (No “hidden agendas”) ..... **HD D N S HS UR**
- 27. Maintain high integrity in all situations ..... **HD D N S HS UR**
- 28. Avoid “playing favorites” ..... **HD D N S HS UR**

**Composure**

- 29. Express confidence in difficult situations ..... **HD D N S HS UR**
- 30. Maintain composure under pressure ..... **HD D N S HS UR**
- 31. Encourage and accept candid criticism ..... **HD D N S HS UR**
- 32. Demonstrate a “can do” attitude (even if others say “it can’t be done”) ..... **HD D N S HS UR**
- 33. Improve as a result of receiving feedback from this questionnaire ..... **HD D N S HS UR**

**Courage**

- 34. Admit mistakes readily ..... **HD D N S HS UR**
- 35. Practice what he or she preaches ..... **HD D N S HS UR**
- 36. Demonstrate courage to stand up for beliefs ..... **HD D N S HS UR**
- 37. Challenge higher level management when appropriate ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**Care For People**

- 38. Realize how he or she affects others ..... **HD D N S HS UR**
- 39. Treat others with respect and dignity ..... **HD D N S HS UR**
- 40. Avoid unnecessary destructive comments about others ..... **HD D N S HS UR**
- 41. Support an appropriate balance between work and non-work life  
for others ..... **HD D N S HS UR**

**CHALLENGE and SUPPORT the ORGANIZATION**

- 42. Represent the organization in a positive manner ..... **HD D N S HS UR**
- 43. Support higher level management ..... **HD D N S HS UR**
- 44. Avoid unnecessary destructive comments about the organization  
and higher level management ..... **HD D N S HS UR**
- 45. Act proud to be a member of the organization ..... **HD D N S HS UR**
- 46. Appropriately challenge decisions he or she thinks are not good for  
the organization ..... **HD D N S HS UR**
- 47. Continuously strive to improve the organization ..... **HD D N S HS UR**

**BUILD TEAMS**

- 48. Actively promote teamwork in his or her area ..... **HD D N S HS UR**
- 49. Share information across organizational boundaries ..... **HD D N S HS UR**
- 50. Establish relationships across a wide range of areas  
throughout the organization ..... **HD D N S HS UR**
- 51. Promote other areas of the organization ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**DEVELOP INDIVIDUALS**

- 52. Provide developmental feedback on performance problems ..... **HD D N S HS UR**
- 53. Give positive feedback for improvement in performance ..... **HD D N S HS UR**
- 54. Provide growth opportunities for others ..... **HD D N S HS UR**
- 55. Support people of all backgrounds to reach their full potential (regardless of race, gender, or other differences) ..... **HD D N S HS UR**
- 56. Provide candid performance feedback in a timely manner ..... **HD D N S HS UR**
- 57. Treat employee mistake as opportunity for instruction ..... **HD D N S HS UR**
- 58. Correct employee’s misconduct ..... **HD D N S HS UR**

**QUALITY and TASK**

- 59. Maintain high quality standards ..... **HD D N S HS UR**
- 60. Communicate the importance of quality ..... **HD D N S HS UR**
- 61. Strive for continuous improvement of relevant work processes ..... **HD D N S HS UR**
- 62. View change as an opportunity ..... **HD D N S HS UR**
- 63. Accept and manage risk ..... **HD D N S HS UR**
- 64. Originate action to complete the task ..... **HD D N S HS UR**
- 65. Accomplish desired results with limited resources when necessary ..... **HD D N S HS UR**
- 66. Keep the right focus ..... **HD D N S HS UR**
- 67. Make the work meaningful and relevant ..... **HD D N S HS UR**
- 68. Explain how the work results in organizational success ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**CUSTOMER/CLIENT Note: a customer or client is any person, internal or external, who receives a service or product. For instance, the following people could be an FBI employee’s customer: Fellow employee, crime victim, accused, witness, defense attorney, and prosecutor.**

- 69. Follow through on commitments made to customers ..... **HD D N S HS UR**
- 70. Avoid sarcastic comments when customers complain ..... **HD D N S HS UR**
- 71. Listen to the ideas of customers for improving work product or services ..... **HD D N S HS UR**
- 72. Consider customers as a top priority ..... **HD D N S HS UR**
- 73. Deal with customer concerns on a timely basis ..... **HD D N S HS UR**
- 74. Assess customer satisfaction levels ..... **HD D N S HS UR**

**VISION**

- 75. Explain the organizational vision to others ..... **HD D N S HS UR**
- 76. Communicate a clear vision for the work group ..... **HD D N S HS UR**
- 77. Motivate others to commit to the vision ..... **HD D N S HS UR**
- 78. Establish a strategy to realize the vision ..... **HD D N S HS UR**
- 79. Implement actions in the work group that are in alignment with the organizational vision ..... **HD D N S HS UR**

**FOSTER CREATIVITY**

- 80. Encourage others to creatively solve problems ..... **HD D N S HS UR**
- 81. Avoid comments that stifle creativity (killer phrases) ..... **HD D N S HS UR**
- 82. Take reasonable risks to implement new ideas ..... **HD D N S HS UR**
- 83. Allow new ideas to be explored and evaluated ..... **HD D N S HS UR**

..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**MANAGE DIVERSITY**

- 84. Provide equal opportunity to compete for assignments or promotions ..... **HD D N S HS UR**
- 85. Create a workplace environment free of intimidation ..... **HD D N S HS UR**
- 86. Create a workplace environment free of hostility ..... **HD D N S HS UR**
- 87. Create a workplace environment free of offensive behavior ..... **HD D N S HS UR**
- 88. Deal with complaints of discrimination effectively ..... **HD D N S HS UR**

**BUILD PUBLIC TRUST** **Note: Government service implies unique responsibilities. Since the clients of Government employees are taxpayers, the services rendered are different from private industry. Taxpayer confidence in government employees is interpreted as “Public Trust.” Also, one who holds a position of “Public Trust” has a fiduciary or trustee type obligation to “The People.”**

**Public Trust**

- 89. Use position only for legitimate personal gain ..... **HD D N S HS UR**
- 90. Ensure others use their position only for legitimate personal gain ..... **HD D N S HS UR**
- 91. Use taxpayers’ funds wisely ..... **HD D N S HS UR**
- 92. Be a team-player when working with other government employees ..... **HD D N S HS UR**
- 93. Assure the public has fair access to services ..... **HD D N S HS UR**

**Integrity**

- 94. Behave the same in private and public life ..... **HD D N S HS UR**
- 95. Exhibit moral behavior consistent with organizational values ..... **HD D N S HS UR**
- 96. Always uphold the Constitution ..... **HD D N S HS UR**
- 97. Always uphold the law ..... **HD D N S HS UR**



..... **ANSWER OPTIONS — CIRCLE ONE**

<b>HD</b>	<b>D</b>	<b>N</b>	<b>S</b>	<b>HS</b>	<b>UR</b>
<i>Highly Dissatisfied</i>	<i>Dissatisfied</i>	<i>Neither Satisfied nor Dissatisfied</i>	<i>Satisfied</i>	<i>Highly Satisfied</i>	<i>Unable to respond</i>

*Consider your effectiveness in the following items.  
How satisfied are others with your ability to . . .*

**PLEASE CONTINUE TO THE NEXT PAGE TO PROVIDE YOUR WRITTEN COMMENTS**

**WRITTEN COMMENTS**

**Please list this person's strengths, opportunities for improvement and general comments below.**

**Strengths:**

---

---

---

---

---

---

---

**Opportunities for Improvement:**

---

---

---

---

---

---

---

**Additional Comments:**

---

---

---

**Thank you for taking the time to complete this form.**



**Developed with Keilty, Goldsmith & Company  
for use in FBI management development programs**

**IMPORTANT**

*Please mail completed form within three days to:*

*NCS Organization Assessment Systems  
4401 West 76th Street  
Edina, MN 55435*

**MSU-WHS: 4.28.04**

# APPENDIX C

*FBI Leadership Commitments  
and Credibility Questionnaire*

Validity and Reliability Analysis

Prepared by:  
Leadership Research Institute  
November 2001

## **FBI Leadership Commitments and Credibility Questionnaire** Validity and Reliability Analyses

This report presents the findings of empirical analyses of the FBI *LEADERSHIP AND COMMITMENTS AND CREDIBILITY* questionnaire. The analyses were conducted to examine the validity and reliability of the questionnaire, and identify items for review and possible revision.

### **Methodology**

#### **Questionnaire**

The questionnaire contains 97 items responded to using a 5-point satisfaction scale (HD, D, N, S, HS) with an additional "Unable to Respond" option. The items are categorized under 11 major dimensions, some of which broken down into smaller dimensions yielding a total of 19 "categories" of items for the analyses in this report. The labels for the categories are listed in Tables 1 and 2.

#### **Data**

The data file contains the ratings of 795 managers (i.e., Self), 4794 Peers, 5480 Direct reports, and 1 uncoded perspective for a total of 11,070 questionnaires.

#### **Analyses**

Confirmatory factor analysis, item analysis, and alpha internal consistency reliability analyses were carried out. These analyses were used to examine the construct validity and reliability of the data obtained from the questionnaire.

### **Findings**

#### **Construct Validity**

During the development of the questionnaire, the literature base and expert review provided *judgmental* support from the content validity of the feedback questionnaire. Now that sufficient ratings are available, it is appropriate to *empirically* exam how the questionnaire is working.

Confirmatory factor analysis was used to examine the construct validity of the categories assessed by the questionnaire. This empirical technique examines the ratings to ascertain if the assignment of items to categories was reasonable. The earlier item/category assignments carried out during the development stage represent a "hypothetical model" that the analysis confirms is reasonable or not for each item.

Tables 1.1 – 1.4 present the item-level factor loadings for each of the 19 categories (factors). These loadings, which are correlations of the item and category are all quite acceptable. We are able to find only five (5) items with loadings less than .70. These items were the following:

Item	Loading
19. Avoid over-leading (too much structure).	.67
21. Perform the technical/functional skills to do the job.	.66
37. Challenge higher level management when appropriate.	.67
44. Avoid unnecessary destructive comments about the organization and higher level management.	.69
46. Appropriately challenge decisions you think are not good for the organization	.67

The high levels of factor loadings suggest that the confirmatory factor analysis “fit” statistics will meet or exceed professional standards (see Bentler, 1990). To put these fit statistics in perspective, readers should note that we are attempting to “confirm” the hypothesized leadership commitment and credibility model specified in the feedback questionnaire. We are, in fact, “testing” the data to ascertain if they fit the model. The most currently popular fit statistics and their obtained fit statistics are listed below:

Statistics	Fit Index
Bentler-Bonnett Nonormed (Tucker-Lewis, TLI)	.88
Comparative Fit Index CFI)	.89
Bollen (IFI)	.89
Root Mean Square Error of Approximation (RMSEA)	.05

In all cases these fit statistics meet professional standards and lend support to the construct validity of the feedback questionnaire.

### Reliability

The item analysis and alpha internal consistency reliability findings are presented in Tables 2.1-2.8. In these tables the response percentages are presented, followed by the item-level means and standard deviations. The next column of data presents the corrected correlation of the item with the set of remaining items in the category. Further to the right is a column of reliability indices for the category if the respective item was deleted from the questionnaire. Finally, the reliability of the data for the overall category is presented.

Examination of the data in the far right column of Tables 2.1-2.8 indicates that the current category reliabilities for the *LEADERSHIP COMMITMENTS AND CREDIBILITY* questionnaire range from .85 to .95. These are considered to be very high levels of reliability and lend support to the adequate sampling of items (i.e., behaviors) during the development of each category, as well as the accuracy of the resulting data.

Since any good instrument can be "fine tuned," we will identify a few items that could be reviewed. The rating of "Unable to Respond" was selected by at least 10% of the raters for items 4, 5, 12, 15, 33, 51-54, 56-58, 69-75, 84, 88, 89-91, 93-94 (see boxed values). Inspection of the boxed values in the first column of Tables 2.1-2.8 indicates that the categories with several items rated "UR" were Develop Individuals, Customer/Client, and Public Trust. In fact, all of the items in the Develop Individuals and Customer/Client categories had "UR" ratings by greater than 10% of the raters. Table 3 contains a breakout of the percent of "UR" ratings by perspective group. As expected, the highest percents are found for the Peer raters for the Develop Individuals category. Readers should note that more than 10% of both Peers and Direct Reports selected the "UR" rating for the Customer/Client items. Developers should review the items boxed in the first column and consider the nature of the behavior being described so that the reasonableness of the "UR" ratings can be considered.

Items with low or high means and low standard deviations sometimes do not correlate well with other items in their respective category. In Table 2.2 Item 21 - Perform the technical/functional skills to do the job - is an example of this situation. Other items with high means and low standard deviations (e.g., item 42) still tend to correlate sufficiently with their respective category. Items that tend to receive high levels of satisfaction from raters can still be judgmentally important to include in the questionnaire, as long as developers realize that they may not contribute much empirically to the assessment.

Continuing in Tables 2.1-2.8, we consider the correlation of each item with the remaining items in the category (Correlation with Category), and the effect of deleting the respective item from the questionnaire. From an empirical sense we note that none of the correlations are considered to be "low". We are simply identifying the correlations that are low relative to the others in the category list (see boxed values). The items associated with these correlations are candidates for review. If they are deleted, the overall category reliability would be the value listed in the column labeled Category Alpha Reliability if Item Deleted. For example, item-19-Avoid over-leading (too much structure) - has a correlation of .65 with the remaining five items defining the Analyzing Others' Readiness category. The overall category reliability is .90. If item 19 was deleted, the reliability would still be .90. If more than one item were deleted from the category, the resulting category reliability would decrease a few points.

Deleting item 21-Perform the technical/functional skills to do the job - from the Competence category would raise the resulting reliability from .85 to .89. While this is the case, a two item category does not sample enough behaviors to adequately support the content validity of the category. Therefore, item 21 should not be deleted. The remaining items associated with boxed correlations (items 34, 37, 41, and 46) do not appear to effect



the level of reliability if they are deleted. Developers will most likely prefer to keep the items in the questionnaire to enhance content coverage and support content validity of the respective category.

### **Recommendations**

The data presented in this report support the validity and reliability of the *LEADERSHIP COMMITMENTS AND CREDIBILITY* questionnaire. While the item and category-level statistics indicate the instrument works well from a psychometric perspective, fine-tuning is always possible. Such revisions can be made based on literature-based judgment and empirical data. This report provides input to the empirically based decisions.

It is recommended that developers consider the items identified for review in this report. In some cases the item could be deleted, and in others revision would be appropriate.

Table 1.1  
**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY**  
 CFA - Factor Loadings

Total  
 N=11,070

Item	STEPS TO EMPOWERMENT			
	Providing Enabling Structure	Supporting / Collaborating	Empowering Others	Analyzing Others Readiness
Provide enabling structure when needed. (1)	.82			
Demonstrate how to do specific tasks when necessary. (2)	.81			
Make certain goals are understood. (3)	.78			
Change from a "boss" to a "coach." (4)	.79			
Provide appropriate orientation on new assignments. (5)	.81			
Provide timely coaching and feedback. (6)	.82			
Collaborate with others in setting goals. (7)		.78		
Encourage others to participate in making decisions. (8)		.81		
Encourage others to discuss their concerns. (9)		.81		
Support others when they need it. (10)		.81		
Motivate others by building their confidence. (11)		.83		
Eliminate unnecessary sign-offs (approvals). (12)			.73	
Delegate authority as well as responsibility. (13)			.88	
Empower others to make decisions. (14)			.91	
Break the work into specific tasks. (15)				.75
Assess the ability of others to do specific tasks. (16)				.80
Assess the motivation of others to do specific tasks. (17)				.81
Ask others what they need to do their job better. (18)				.74
Avoid over-leading (too much structure). (19)				<b>.67</b>
Avoid under-leading (too little structure). (20)				.73

Continued...

Table 1.2  
**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY**  
 CFA - Factor Loadings

Items	CREDIBILITY				
	Competence	Character	Composure	Courage	Care for People
Perform the technical/functional skills to do the job. (21)	.66				
Manage people. (22)	.88				
Manage conflict effectively. (23)	.81	.78			
Ensure honest reporting of results. (24)		.76			
Follow through on commitments. (25)		.84			
Communicate honestly with others (No "hidden agendas"). (26)		.86			
Maintain high integrity in all situations. (27)		.75			
Avoid "playing favorites." (28)			.77		
Express confidence in difficult situations. (29)			.75		
Maintain composure under pressure. (30)			.74		
Encourage and accept candid criticism. (31)			.74		
Demonstrate a "can do" attitude (even if others say "it can't be done.") (32)				.73	
Improve as a result of receiving feedback from this questionnaire. (33)				.84	
Admit mistakes readily. (34)				.77	
Practice what you preach. (35)				.67	
Demonstrate courage to stand up for beliefs. (36)					.84
Challenge higher level management when appropriate. (37)					.89
Realize how he or she affects others. (38)					.82
Treat others with respect and dignity. (39)					.74
Avoid unnecessary destructive comments about others. (40)					
Support an appropriate balance between work and non-work life for others. (41)					

Continued

Table 1.3  
 FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
 CFA - Factor Loadings  
 Total  
 N=11,070

Item	CHALLENGE and SUPPORT the ORGANIZATION	BUILD TEAMS	DEVELOP INDIVIDUALS	QUALITY and TASKS
Represent the organization in a positive manner. (42)	.77			
Support higher level management. (43)	.72			
Avoid unnecessary destructive comments about the organization and higher level management. (44)	.69			
Act proud to be a member of the organization. (45)	.73			
Appropriately challenge decisions you think are not good for organization. (46)	.67			
Continuously strive to improve the organization. (47)	.79	.79		
Actively promote teamwork in your area. (48)		.79		
Share information across organizational boundaries. (49)		.64		
Establish relationships across a wide range of areas throughout the organization. (50)			.83	
Promote other areas of the organization. (51)			.83	
Provide developmental feedback on performance problems. (52)			.79	
Give positive feedback for improvement in performance. (53)			.71	
Provide growth opportunities for others. (54)			.82	
Support people of all backgrounds to reach their full potential (regardless of race, gender, or other differences). (55)			.81	
Provide careful performance feedback in a timely manner. (56)			.72	.75
Treat employee mistake as opportunity for instruction. (57)				.79
Correct employee's misconduct. (58)				.81
Maintain high quality standards. (59)				.75
Communicate the importance of quality. (60)				
Strive for continuous improvement of relevant work processes. (61)				
View change as an opportunity. (62)				

Continued

Table 1.4  
**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY**  
 CFA - Factor Loadings

Item	Total		QUALITY and TASKS	CUSTOMER/ CLIENT	VISION	FOSTER CREATIVITY
	N=11,070					
Accept and manage risk. (63)	.79					
Originate action to complete the task. (64)	.81					
Accomplished desired results with limited resources when necessary. (65)	.78					
Keep the right focus. (66)	.80					
Make the work meaningful and relevant. (67)	.82					
Explain how the work results in organizational success. (68)	.81	.75				
Follow through on commitments made to customers. (69)		.72				
Avoid sarcastic remarks when customers complain. (70)		.80				
Listen to the ideas of customers for improving work product or services. (71)		.83				
Consider customers as a top priority. (72)		.83				
Deal with customer concerns on a timely basis. (73)		.84				
Assess customer satisfaction levels. (74)					.83	
Explain the organizational vision to others. (75)					.88	
Communicate a clear vision for the work group. (76)					.89	
Motivate others to commit to the vision. (77)					.91	
Establish a strategy to realize the vision. (78)					.89	
Implement actions in the work group that are in alignment with the organizational vision. (79)						.83
Encourage others to creatively solve problems. (80)						.80
Avoid comments that stifle creativity (killer phrases). (81)						.85
Take reasonable risks to implement new ideas. (82)						.86
Allow new ideas to be explored and evaluated. (83)						

Continued...

Table 1.5

FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
CFA - Factor Loadings

Items	Total N=11,070	BUILD PUBLIC TRUST	
		MANAGE DIVERSITY	Public Trust Integrity
Provides equal opportunity to compete for assignments or promotions. (84)	.73		
Create a workplace environment free of intimidation. (85)	.88		
Create a workplace environment free of hostility. (86)	.89		
Create a workplace environment free of offensive behavior. (87)	.85		
Deal with complaints of discrimination effectively. (88)	.79		
Use position only for legitimate personal gains. (89)	.90		
Ensure others use their position only for legitimately personal gains. (90)	.92		
Use taxpayers' funds wisely. (91)	.82		
Be a team-player when working with other government employees. (92)	.74		
Assure the public has fair access to services. (93)	.79		.76
Behave the same in private and in public life. (94)			.81
Exhibit moral behavior consistent with organizational values. (95)			.94
Always uphold the Constitution. (96)			.93
Always uphold the law. (97)			

Table 2.1

**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
ITEM ANALYSIS AND RELIABILITY DATA**

N=11,070

Categories/Items	Response Percentages <sup>1</sup>					Mean <sup>2</sup>	Standard <sup>2</sup> Deviation	Correlation with Category	Category Alpha Reliability If Item Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>STEPS TO EMPOWERMENT</b>										
<b>Providing Enabling Structure</b>										
Provide enabling structure when needed. (1)	10	<1	3	7	38	41	.80	.83	.92	.93
Demonstrate how to do specific tasks when necessary. (2)	9	<1	3	9	39	40	.81	.82	.92	
Make certain goals are understood. (3)	7	<1	2	6	38	46	.79	.78	.93	
Change from a "boss" to a "coach." (4)	13	1	4	11	31	39	.92	.78	.93	
Provide appropriate orientation on new assignments. (5)	11	<1	3	9	39	37	.84	.82	.92	
Provide timely coaching and feedback. (6)	10	1	4	10	37	38	.90	.82	.92	
<b>Supporting/Collaborating</b>										
Collaborate with others in setting goals. (7)	5	1	4	9	37	45	.86	.76	.91	.92
Encourage others to participate in making decisions. (8)	5	1	4	9	34	47	.90	.82	.70	
Encourage others to discuss their concerns. (9)	4	1	4	8	33	51	.87	.81	.68	
Support others when they need it. (10)	3	2	3	7	30	55	.89	.80	.90	
Motivate others by building their confidence. (11)	6	2	5	11	33	43	.96	.80	.90	
<b>Empowering Others</b>										
Eliminate unnecessary sign-offs (approvals). (12)	18	1	4	12	33	33	.92	.71	.90	
Delegate authority as well as responsibility. (13)	9	1	4	8	35	42	.91	.83	.80	

NOTE: It is recommended that items with boxed entries be examined for possible revision or deletion.

<sup>1</sup> Rating Scale: UR=Unable to Respond, 1=Highly Dissatisfied, 2=Dissatisfied, 3=Neither Satisfied nor Dissatisfied, 4=Satisfied, 5=Highly Satisfied.<sup>2</sup> Mean and standard deviation calculated on complete sets of data for items within each category.

Continued...

Table 2.2  
**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
 ITEM ANALYSIS AND RELIABILITY DATA**  
 N=11,070

Categories/Items	Response Percentages <sup>a</sup>					Mean <sup>b</sup>	Standard <sup>c</sup> Deviation	Correlation with Category	Category Alpha Reliability If Items Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>Empowering Others (continued)</b>										
Empower others to make decisions. (14)	9	1	4	9	36	40	4.21	.91	.81	.90
<b>Analyzing Others' Readiness</b>										
Break the work into specific tasks. (15)	12	<1	2	9	41	36	4.24	.76	.88	
Assess the ability of others to do specific tasks. (16)	9	<1	3	7	40	40	4.28	.80	.87	
Assess the motivation of others to do specific tasks. (17)	10	<1	3	10	39	37	4.21	.83	.87	
Ask others what they need to do their job better. (18)	9	1	4	10	35	42	4.24	.87	.88	
Avoid over-leading (too much structure). (19)	9	2	5	12	35	37	4.12	.96	.90	
Avoid under-leading (too little structure). (20)	10	1	4	14	37	34	4.12	.89	.88	
<b>CREDIBILITY</b>										
<b>Competence</b>										.85
Perform the technical/functional skills to do the job. (21)	4	<1	2	5	29	59	4.48	.79	.63	.89
Manage people. (22)	3	2	5	9	35	45	4.21	.97	.81	.70
Manage conflict effectively. (23)	6	3	7	12	35	38	4.07	1.02	.77	.76
<b>Character</b>										.92
Ensure honest reporting of results. (24)	4	<1	2	5	26	63	4.56	.73	.78	.90
Follow through on commitments. (25)	3	<1	3	6	29	59	4.47	.81	.76	.90
Communicate honestly with others (No "hidden agendas"). (26)	3	2	5	8	26	57	4.36	.95	.84	.89

Continued...



Table 2.3

**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
ITEM ANALYSIS AND RELIABILITY DATA**

N=11,070

Categories/Items	Response Percentages <sup>a</sup>					Mean <sup>b</sup>	Standard <sup>c</sup> Deviation	Correlation with Category	Category Alpha Reliability If Item Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>Character</b> (continued)										
Maintain high integrity in all situations. (27)	2	1	2	6	24	65	4.53	.85	.89	.90
Avoid "playing favorites." (28)	6	2	5	10	30	47	4.23	.75	.91	
<b>Composure</b>										
Express confidence in difficult situations. (29)	3	<1	2	7	34	53	4.41	.77	.87	.88
Maintain composure under pressure. (30)	4	<1	3	8	34	51	4.36	.77	.87	
Encourage and accept candid criticism. (31)	9	2	5	12	36	37	4.15	.75	.88	.88
Demonstrate a "can do" attitude (even if others say "it can't be done.") (32)	4	<1	2	7	33	54	4.45	.72	.88	
Improve as a result of receiving feedback from this questionnaire. (33)	17	1	2	9	30	41	4.30	.73	.88	.87
<b>Courage</b>										
Admit mistakes readily. (34)	8	2	4	10	38	38	4.18	.68	.85	.89
Practice what you preach. (35)	5	1	3	8	34	49	4.33	.77	.82	
Demonstrate courage to stand up for beliefs. (36)	4	<1	2	7	39	56	4.44	.78	.82	.86
Challenge higher level management when appropriate. (37)	8	2	5	9	30	46	4.24	.68	.86	
<b>Care for People</b>										
Realize how he or she affects others. (38)	5	2	6	11	38	39	4.14	.76	.86	.83
Treat others with respect and dignity. (39)	1	1	3	6	29	60	4.46	.83	.83	

Continued...

Table 2.4

**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
ITEM ANALYSIS AND RELIABILITY DATA**

N=11,070

Categories/Items	Response Percentages <sup>a</sup>					Mean <sup>b</sup>	Standard <sup>c</sup> Deviation	Correlation with Category	Category Alpha Reliability If Item Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>Care for People (continued)</b>										
Avoid unnecessary destructive comments about others. (40)	3	2	4	7	32	53	.89	.77	.85	.89
Support an appropriate balance between work and non-work life for others. (41)	10	<1	2	7	31	50	.78	.68	.89	
<b>CHALLENGE and SUPPORT the ORGANIZATION</b>										
Represent the organization in a positive manner. (42)	1	<1	1	3	25	70	.61	.73	.86	.90
Support higher level management. (43)	4	<1	<1	4	33	58	.64	.73	.86	
Avoid unnecessary destructive comments about the organization and higher level management. (44)	3	<1	1	6	33	57	.68	.70	.87	
Act proud to be a member of the organization. (45)	2	<1	<1	3	24	71	.54	.73	.87	
Appropriately challenge decisions you think are not good for the organization. (46)	8	<1	2	8	31	50	.80	.62	.89	
Continuously strive to improve the organization. (47)	4	<1	1	7	31	57	.69	.75	.86	
<b>BUILD TEAMS</b>										
Actively promote teamwork in your area. (48)	6	1	3	8	33	48	.85	.74	.89	
Share information across organizational boundaries. (49)	7	1	3	7	35	47	.83	.79	.87	
Establish relationships across a wide range of areas throughout the organization. (50)	8	<1	2	8	34	48	.80	.82	.86	
Promote other areas of the organization. (51)	12	<1	2	11	36	38	.80	.78	.87	Continued...

Table 2.5

**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
ITEM ANALYSIS AND RELIABILITY DATA**

N=11,070

Categories/Items	Response Percentages <sup>a</sup>					Mean <sup>b</sup>	Standard <sup>c</sup> Deviation	Correlation with Category	Category Alpha Reliability If Item Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>DEVELOP INDIVIDUALS</b>										
Provide developmental feedback on performance problems. (52)	18	<1	3	9	37	32	4.18	.84	.83	.92
Give positive feedback for improvement in performance. (53)	16	<1	3	8	36	37	4.26	.82	.83	.92
Provide growth opportunities for others. (54)	15	1	2	9	34	39	4.27	.84	.79	.92
Support people of all backgrounds to reach their full potential (regardless of race, gender, or other differences). (55)	9	<1	1	5	26	59	4.52	.77	.73	.93
Provide candid performance feedback in a timely manner. (56)	15	<1	3	8	35	38	4.25	.84	.82	.92
Treat employee mistake as opportunity for instruction. (57)	18	1	3	10	36	33	4.19	.84	.80	.92
Correct employee's misconduct. (58)	21	1	3	10	34	30	4.14	.88	.72	.93
<b>QUALITY and TASK</b>										
Maintain high quality standards. (59)	1	<1	<1	4	31	62	4.58	.65	.77	.95
Communicate the importance of quality. (60)	4	<1	1	5	34	55	4.51	.67	.79	.95
Strive for continuous improvement of relevant work processes. (61)	5	<1	1	6	36	51	4.46	.69	.82	.95
View change as an opportunity. (62)	8	<1	2	9	38	42	4.33	.74	.76	.95
Accept and manage risk. (63)	8	<1	3	9	36	44	4.33	.80	.80	.95
Originate action to complete the task. (64)	5	<1	2	7	37	49	4.42	.72	.81	.95
Accomplished desired results with limited resources when necessary. (65)	6	<1	1	6	36	51	4.46	.69	.80	.95

Continued...

Table 2.6

**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
ITEM ANALYSIS AND RELIABILITY DATA**

N=11,070

Categories/Items	Response Percentages <sup>a</sup>					Mean <sup>b</sup>	Standard <sup>c</sup> Deviation	Correlation with Category	Category Alpha Reliability If Item Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>QUALITY and TASK (continued)</b>										
Keep the right focus. (66)	4	<1	2	7	37	49	.77	.81	.95	
Make the work meaningful and relevant. (67)	7	<1	2	9	39	43	.90	.82	.95	
Explain how the work results in organizational success. (68)	10	<1	2	11	37	40	.79	.80	.95	
<b>CUSTOMER/CLIENT</b>										
Follow through on commitments made to customers. (69)	12	<1	2	5	34	48	.73	.75	.92	.93
Avoid sarcastic remarks when customers complain. (70)	13	<1	3	8	33	43	.81	.72	.92	
Listen to the ideas of customers for improving work product or services. (71)	14	<1	2	6	35	42	.75	.80	.91	
Consider customers as a top priority. (72)	12	<1	2	8	34	45	.75	.83	.91	
Deal with customer concerns on a timely basis. (73)	12	<1	2	6	35	45	.75	.83	.91	
Assess customer satisfaction levels. (74)	18	<1	2	10	35	35	.78	.82	.91	.95
<b>VISION</b>										
Explain the organizational vision to others. (75)	12	<1	2	10	38	39	.77	.83	.95	
Communicate a clear vision for the work group. (76)	9	<1	3	9	37	41	.82	.88	.94	
Motivate others to commit to the vision. (77)	1	1	3	12	38	35	.86	.87	.77	
Establish a strategy to realize the vision. (78)	11	<1	3	12	37	37	.84	.90	.94	
Implement actions in the work group that are in alignment with the organizational vision. (79)	11	<1	2	10	38	38	.79	.87	.94	Continued...

Table 2.7

**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
ITEM ANALYSIS AND RELIABILITY DATA**

N=11,070

Categories/Items	Response Percentages <sup>1</sup>					Mean <sup>3</sup>	Standard <sup>2</sup> Deviation	Correlation with Category	Category Alpha Reliability If Item Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>FOSTER CREATIVITY</b>										
Encourage others to creatively solve problems. (80)	9	<1	2	8	38	42	.79	.80	.89	.91
Avoid comments that stifle creativity (killer phrases). (81)	8	1	3	9	35	43	.85	.77	.90	
Take reasonable risks to implement new ideas. (82)	9	<1	3	9	38	40	.82	.81	.89	
Allow new ideas to be explored and evaluated. (83)	7	<1	2	8	37	45	.79	.84	.88	
<b>MANAGE DIVERSITY</b>										
Provide equal opportunity to compete for assignments or promotions. (84)	13	1	3	6	28	49	.81	.76	.94	
Create a workplace environment free of intimidation. (85)	7	2	3	6	28	55	.86	.87	.92	
Create a workplace environment free of hostility. (86)	6	2	4	7	28	53	.88	.88	.91	
Create a workplace environment free of offensive behavior. (87)	6	<1	2	5	30	56	.76	.85	.92	
Deal with complaints of discrimination effectively. (88)	24	<1	1	5	23	46	.76	.82	.93	.92
<b>BUILD PUBLIC TRUST</b>										
<b>Public Trust</b>										
Use position only for legitimate personal gain. (89)	17	<1	<1	4	21	56	.68	.81	.90	
Ensure others use their position only for legitimate personal gain. (90)	18	<1	<1	5	25	52	.68	.84	.89	
Use taxpayers' funds wisely. (91)	12	<1	<1	4	26	57	.63	.81	.90	
Be a team-player when working with other government employees. (92)	7	<1	1	4	28	60	.68	.72	.91	
Assure the public has fair access to services. (93)	19	<1	<1	4	26	49	.62	.79	.90	Continued...

Table 2.8

**FBI LEADERSHIP COMMITMENTS AND CREDIBILITY  
ITEM ANALYSIS AND RELIABILITY DATA**

N=11,070

Categories/Items	Response Percentages <sup>1</sup>					Mean <sup>2</sup>	Standard <sup>3</sup> Deviation	Correlation with Category	Category Alpha Reliability If Item Deleted	Overall Value Alpha Reliability
	UR	1	2	3	4					
<b>Integrity</b>										.93
Behave the same in private and in public life. (94)	23	<1	<1	3	18	55	.61	.81	.92	
Exhibit moral behavior consistent with organizational values. (95)	6	<1	<1	3	22	68	.57	.85	.90	
Always uphold the Constitution. (96)	9	<1	<1	2	18	71	.47	.85	.90	
Always uphold the law. (97)	8	<1	<1	2	17	71	.46	.84	.90	

**Table 3**  
 Percent of Raters Selecting "Unable to Respond"  
 by Perspective Group

Category/Item	Manager N=795	Peer N=4,794	Direct Report N=5,480
<b>Develop Individuals</b>			
Provide developmental feedback on performance problems. (52)	2	32	9
Give positive feedback for improvement in performance. (53)	1	30	6
Provide growth opportunities for others. (54)	4	26	7
Support people of all backgrounds to reach their full potential (regardless of race, gender, or other differences). (55)	1	15	4
Provide candid performance feedback in a timely manner. (56)	1	30	5
Treat employee mistake as opportunity for instruction. (57)	3	31	9
Correct employee's misconduct. (58)	13	30	17
<b>Customer/Client</b>			
Follow through on commitments made to customers. (69)	1	13	13
Avoid sarcastic remarks when customers complain. (70)	2	15	13
Listen to the ideas of customers for improving work product or services. (71)	2	15	15
Consider customers as a top priority. (72)	2	13	12
Deal with customer concerns on a timely basis. (73)	1	14	12
Assess customer satisfaction levels. (74)	3	20	18

# APPENDIX D





Leadership and  
Management Science Unit  
FBI Academy  
Quantico, Virginia 22135  
(703) 632-3154

June 8, 2001

Dear EDI Participant:

Congratulations on being selected by the SAMMS Board to attend the next Executive Development Institute (EDI) II, 08/13-24/2001. The agenda for the two weeks is enclosed.

#### **COURSE**

EDI II is designed for your personal development as a leader in the FBI. It will stretch your mind in new directions by introducing unique ideas, knowledge, and skills. The intent is to provide you with the ability and confidence to accept new responsibilities. Think of EDI as an important step in your continuous self-development.

#### **TRAVEL ARRANGEMENTS**

If driving, arrange to arrive no later than 7:30 AM, Tuesday, 08/14/2001. For those coming by plane, a bus will be scheduled to leave National Airport on Monday, 08/13/2001, at 7 p.m.

*Your Travel Authorization Number is TR51-AG-010212.*

#### **WILL YOU ATTEND?**

The first thing you should do is determine whether or not you will attend. If you cannot, please call Lisa Bush, Executive Programs Administrative Assistant, at Quantico, extension 3154. Then, prepare a communication to EDSP, Administrative Services Division, as to why you cannot participate. Also, return the 15 forms to Mrs. Bush, Leadership and Management Science Unit, Quantico, noting they are from you. If you have any questions or concerns, please feel free to contact her. She will assist you.

#### **DO THIS NOW**

If you can attend, you should **immediately** arrange to complete the following instructions with the *Leadership Commitments and Credibility Inventory System* packet. Please meet the deadlines for mailing the completed forms.

The goal is to have the "Other Forms" filled out by seven peers and by seven direct reports. Peers are defined as people organizationally on your level, and direct reports are people you supervise. Having six of one group and eight of the other, or even five/nine is permissible. You complete and mail the "Self Form." I recommend you give the assignment for distribution of the "Other Forms" to your assistant with instructions to distribute them randomly (so as to enhance confidentiality and scope of feedback).

Assure your name is filled in on each instrument and check (✓) whether evaluator is a peer or direct report. Have an assistant place enclosed labels to Questar on 10 1/2 x 12 inch envelopes. Thus, the forms can be returned directly by those providing feedback to assure confidentiality. Your assistant should follow up to ensure forms are completed and mailed.

The feedback will be provided to you by Dr. John Greene, an outside consultant, on a confidential basis. The feedback is for your self-development only.

#### **ATTIRE**

Classroom attire is normally casual. However, during media presentations and some dinners, you are requested to wear business attire. It makes the feedback more meaningful for your public speaking and press release TV appearances. Business attire is also traditionally worn during graduation and our scheduled evening meals. Further information about dress is explained in the enclosed General Information sheet.

#### **TELEPHONE--ADDRESS--FAX**

To help you communicate with the outside world, I have enclosed two helpful sheets. One is to help you place calls. The other is to leave with people needing to communicate with you. You will be able to call out from your rooms in the Jefferson Dorm. Instruct people to leave messages on your dorm room VoiceMail. Please make plans to ensure telephone calls will not be disruptive to your attendance in class.

#### **ARRIVAL**

Your EDI badge, room assignment, parking permit, and access card will be provided at the reception desk. Our classroom will be in the Lyceum, Jefferson Building.

#### **FIRST MORNING**

We will begin with a breakfast at 7:30 a.m. on Tuesday, August 14, 2001, in the Lyceum.

I look forward to the next EDI session and hope to make your stay here informative and comfortable. Call or e-mail Dawn Leonard if your name or assignment is incorrect.

Sincerely,

Kevin R. Cornelius  
(703) 632-3135

#### **Enclosures and Attachments (5)**

1. EDI Tentative Schedule
2. General Information for Academy Training
3. Phone Information
4. I Can be Reached
5. *Leadership Commitments and Credibility Inventory System* packet

## Leadership Commitments and Credibility Inventory System

The Leadership Commitments and Credibility Inventory System (LCCIS) is a feedback system designed to provide comment regarding performance to the rated individual from their subordinates and peers. It is tool designed for *self-development* only and results will be distributed only to you, the participant. Dr. John Greene, who is a consultant for the FBI with a ten-year relationship, will be here in September to discuss the results obtained from this instrument.

### *Goal*

The goal is to have the "Other Forms" filled out by seven peers and by seven direct reports. Peers are defined as people organizationally on your level, and direct reports are people you supervise. Having six of one group and eight of the other, or even five/nine is permissible. You complete and mail the "Self Form." I recommend you give the assignment for distribution of the "Other Forms" to your assistant with instructions to distribute them randomly (so as to enhance confidentiality and scope of feedback).

### *Action Steps*

Assure your **name is filled in on each instrument and check (✓) whether evaluator is a peer or a direct report.** Have an assistant place enclosed labels to Questar on 10 ½ x 12-inch envelopes. Thus, the forms can be returned directly by those providing feedback to assure confidentiality. Your assistant should follow up to ensure forms are completed and mailed.

The feedback will be provided to you when you return in September. Dr. Greene will be here then to discuss leadership and performance issues and assist in questions regarding self-development.

# APPENDIX E

## APPENDIX E

### RELIABILITY ANALYSIS - ALPHA SCALES (Time One)

#### STEPS TO EMPOWERMENT

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 20      Alpha = .9629

#### CREDIBILITY

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 21      Alpha = .9664

#### CHALLENGE and SUPPORT the ORGANIZATION

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 6      Alpha = .8872

#### BUILD TEAMS

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 4      Alpha = .8935

#### DEVELOP INDIVIDUALS

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 7      Alpha = .8442

#### QUALITY and TASK

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 10      Alpha = .9498

#### CUSTOMER/CLIENT

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 6      Alpha = .9206

#### VISION

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 5      Alpha = .9453

#### FOSTER CREATIVITY

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 4      Alpha = .9115

#### MANAGE DIVERSITY

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 5      Alpha = .9134

#### BUILD PUBLIC TRUST

Reliability Coefficients  
N of Cases = 1014.0      N of Items = 9      Alpha = .9188

# APPENDIX F

Analysis of Covariance Testing: Hypothesis 3

**Between-Subjects Factors**

		N
RESPTYPE	M	73
	P	73

M = self

P = peer

N = 73 in all cases for both self and peers

**Overall Measurement**

**Tests of Between-Subjects Effects**

Dependent Variable: ALL2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.057 <sup>a</sup>	2	3.029	37.650	.000
Intercept	2.128	1	2.128	26.456	.000
ALL1	5.712	1	5.712	71.008	.000
RESPTYPE	8.210E-02	1	8.210E-02	1.021	.314
Error	11.503	143	8.044E-02		
Total	2955.190	146			
Corrected Total	17.560	145			

a. R Squared = .345 (Adjusted R Squared = .336)

**Steps to Empowerment**

**Tests of Between-Subjects Effects**

Dependent Variable: STOE2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.623 <sup>a</sup>	2	2.812	29.114	.000
Intercept	4.129	1	4.129	42.751	.000
STOE1	5.268	1	5.268	54.551	.000
RESPTYPE	6.128E-02	1	6.128E-02	.635	.427
Error	13.810	143	9.657E-02		
Total	2837.200	146			
Corrected Total	19.433	145			

a. R Squared = .289 (Adjusted R Squared = .279)

## Credibility

### Tests of Between-Subjects Effects

Dependent Variable: CRED2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8.283 <sup>a</sup>	2	4.141	42.205	.000
Intercept	2.189	1	2.189	22.304	.000
CRED1	8.098	1	8.098	82.522	.000
RESPTYPE	6.569E-02	1	6.569E-02	.669	.415
Error	14.032	143	9.813E-02		
Total	2932.200	146			
Corrected Total	22.315	145			

a. R Squared = .371 (Adjusted R Squared = .362)

## Challenge and Support the Organization

### Tests of Between-Subjects Effects

Dependent Variable: CSO2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.606 <sup>a</sup>	2	3.303	29.410	.000
Intercept	2.838	1	2.838	25.272	.000
CSO1	5.807	1	5.807	51.707	.000
RESPTYPE	.116	1	.116	1.032	.311
Error	16.061	143	.112		
Total	3091.820	146			
Corrected Total	22.667	145			

a. R Squared = .291 (Adjusted R Squared = .282)

## Build Teams

### Tests of Between-Subjects Effects

Dependent Variable: BT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.077 <sup>a</sup>	2	3.039	15.589	.000
Intercept	9.660	1	9.660	49.560	.000
BT1	5.814	1	5.814	29.828	.000
RESPTYPE	8.069E-03	1	8.069E-03	.041	.839
Error	27.873	143	.195		
Total	2954.560	146			
Corrected Total	33.950	145			

a. R Squared = .179 (Adjusted R Squared = .168)



## Develop Individuals

### Tests of Between-Subjects Effects

Dependent Variable: DI2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.072 <sup>a</sup>	2	1.036	6.432	.002
Intercept	13.338	1	13.338	82.819	.000
DI1	1.857	1	1.857	11.531	.001
RESPTYPE	2.639E-02	1	2.639E-02	.164	.686
Error	23.030	143	.161		
Total	2910.040	146			
Corrected Total	25.102	145			

a. R Squared = .083 (Adjusted R Squared = .070)

## Quality and Task

### Tests of Between-Subjects Effects

Dependent Variable: QT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.209 <sup>a</sup>	2	2.605	19.370	.000
Intercept	5.662	1	5.662	42.103	.000
QT1	4.679	1	4.679	34.795	.000
RESPTYPE	.190	1	.190	1.413	.237
Error	19.230	143	.134		
Total	2984.540	146			
Corrected Total	24.439	145			

a. R Squared = .213 (Adjusted R Squared = .202)

## Customer/Client

### Tests of Between-Subjects Effects

Dependent Variable: CC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.650 <sup>a</sup>	2	2.825	18.765	.000
Intercept	8.323	1	8.323	55.283	.000
CC1	4.924	1	4.924	32.704	.000
RESPTYPE	.604	1	.604	4.009	.047
Error	21.529	143	.151		
Total	2872.250	146			
Corrected Total	27.180	145			

a. R Squared = .208 (Adjusted R Squared = .197)

## Vision

### Tests of Between-Subjects Effects

Dependent Variable: V2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.102 <sup>a</sup>	2	4.551	27.892	.000
Intercept	6.439	1	6.439	39.462	.000
V1	8.288	1	8.288	50.796	.000
RESPTYPE	.219	1	.219	1.343	.248
Error	23.332	143	.163		
Total	2838.790	146			
Corrected Total	32.434	145			

a. R Squared = .281 (Adjusted R Squared = .271)

## Foster Creativity

### Tests of Between-Subjects Effects

Dependent Variable: FC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.089 <sup>a</sup>	2	3.045	19.057	.000
Intercept	11.376	1	11.376	71.205	.000
FC1	6.072	1	6.072	38.004	.000
RESPTYPE	4.690E-02	1	4.690E-02	.294	.589
Error	22.847	143	.160		
Total	2910.320	146			
Corrected Total	28.937	145			

a. R Squared = .210 (Adjusted R Squared = .199)

## Manage Diversity

### Tests of Between-Subjects Effects

Dependent Variable: MD2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.974 <sup>a</sup>	2	2.487	21.640	.000
Intercept	4.779	1	4.779	41.585	.000
MD1	4.967	1	4.967	43.220	.000
RESPTYPE	1.996E-02	1	1.996E-02	.174	.678
Error	16.436	143	.115		
Total	3216.540	146			
Corrected Total	21.410	145			

a. R Squared = .232 (Adjusted R Squared = .222)

## Build Public Trust

### Tests of Between-Subjects Effects

Dependent Variable: BPT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.752 <sup>a</sup>	2	1.376	12.872	.000
Intercept	3.372	1	3.372	31.537	.000
PBT1	2.514	1	2.514	23.513	.000
RESPTYPE	.213	1	.213	1.993	.160
Error	15.288	143	.107		
Total	3262.950	146			
Corrected Total	18.040	145			

a. R Squared = .153 (Adjusted R Squared = .141)

# APPENDIX G

Analysis of Covariance Testing: Hypothesis 4

**Between-Subjects Factors**

		N
RESPTYPE	M	73
	S	70

M = self

S = subordinates

N = 73 for self and 70 for subordinates

**Overall Measurement**

**Tests of Between-Subjects Effects**

Dependent Variable: ALL2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.086 <sup>a</sup>	2	2.543	20.380	.000
Intercept	4.110	1	4.110	32.938	.000
ALL1	4.665	1	4.665	37.393	.000
RESPTYPE	.132	1	.132	1.057	.306
Error	17.468	140	.125		
Total	2770.830	143			
Corrected Total	22.553	142			

a. R Squared = .225 (Adjusted R Squared = .214)

**Steps to Empowerment**

**Tests of Between-Subjects Effects**

Dependent Variable: STOE2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.265 <sup>a</sup>	2	2.633	20.089	.000
Intercept	6.162	1	6.162	47.022	.000
STOE1	4.880	1	4.880	37.238	.000
RESPTYPE	.191	1	.191	1.454	.230
Error	18.348	140	.131		
Total	2659.090	143			
Corrected Total	23.613	142			

a. R Squared = .223 (Adjusted R Squared = .212)

## Credibility

### Tests of Between-Subjects Effects

Dependent Variable: CRED2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8.289 <sup>a</sup>	2	4.145	27.782	.000
Intercept	4.793	1	4.793	32.126	.000
CRED1	7.452	1	7.452	49.953	.000
RESPTYPE	.129	1	.129	.868	.353
Error	20.886	140	.149		
Total	2739.880	143			
Corrected Total	29.175	142			

a. R Squared = .284 (Adjusted R Squared = .274)

## Challenge and Support the Organization

### Tests of Between-Subjects Effects

Dependent Variable: CSO2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3.681 <sup>a</sup>	2	1.840	12.000	.000
Intercept	5.704	1	5.704	37.192	.000
CSO1	3.678	1	3.678	23.984	.000
RESPTYPE	5.783E-03	1	5.783E-03	.038	.846
Error	21.472	140	.153		
Total	2929.910	143			
Corrected Total	25.153	142			

a. R Squared = .146 (Adjusted R Squared = .134)

## Build Teams

### Tests of Between-Subjects Effects

Dependent Variable: BT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.306 <sup>a</sup>	2	2.153	9.189	.000
Intercept	14.926	1	14.926	63.700	.000
BT1	3.647	1	3.647	15.563	.000
RESPTYPE	.505	1	.505	2.156	.144
Error	32.805	140	.234		
Total	2760.020	143			
Corrected Total	37.111	142			

a. R Squared = .116 (Adjusted R Squared = .103)

## Develop Individuals

### Tests of Between-Subjects Effects

Dependent Variable: DI2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.630 <sup>a</sup>	2	1.315	7.487	.001
Intercept	12.652	1	12.652	72.046	.000
DI1	2.211	1	2.211	12.589	.001
RESPTYPE	.515	1	.515	2.935	.089
Error	24.586	140	.176		
Total	2737.920	143			
Corrected Total	27.215	142			

a. R Squared = .097 (Adjusted R Squared = .084)

## Quality and Task

### Tests of Between-Subjects Effects

Dependent Variable: QT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3.750 <sup>a</sup>	2	1.875	10.618	.000
Intercept	8.054	1	8.054	45.612	.000
QT1	3.507	1	3.507	19.860	.000
RESPTYPE	.142	1	.142	.805	.371
Error	24.720	140	.177		
Total	2799.590	143			
Corrected Total	28.469	142			

a. R Squared = .132 (Adjusted R Squared = .119)

## Customer/Client

### Tests of Between-Subjects Effects

Dependent Variable: CC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.725 <sup>a</sup>	2	.862	4.824	.009
Intercept	13.071	1	13.071	73.111	.000
CC1	1.710	1	1.710	9.564	.002
RESPTYPE	5.516E-02	1	5.516E-02	.309	.579
Error	25.030	140	.179		
Total	2737.460	143			
Corrected Total	26.755	142			

a. R Squared = .064 (Adjusted R Squared = .051)

## Vision

### Tests of Between-Subjects Effects

Dependent Variable: V2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.607 <sup>a</sup>	2	3.803	17.864	.000
Intercept	7.391	1	7.391	34.713	.000
V1	7.419	1	7.419	34.846	.000
RESPTYPE	.119	1	.119	.561	.455
Error	29.808	140	.213		
Total	2649.760	143			
Corrected Total	37.414	142			

a. R Squared = .203 (Adjusted R Squared = .192)

## Foster Creativity

### Tests of Between-Subjects Effects

Dependent Variable: FC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.344 <sup>a</sup>	2	3.672	19.349	.000
Intercept	11.033	1	11.033	58.136	.000
FC1	6.477	1	6.477	34.127	.000
RESPTYPE	.438	1	.438	2.307	.131
Error	26.570	140	.190		
Total	2746.360	143			
Corrected Total	33.914	142			

a. R Squared = .217 (Adjusted R Squared = .205)

## Manage Diversity

### Tests of Between-Subjects Effects

Dependent Variable: MD2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8.653 <sup>a</sup>	2	4.327	23.987	.000
Intercept	6.722	1	6.722	37.270	.000
MD1	6.509	1	6.509	36.089	.000
RESPTYPE	.294	1	.294	1.629	.204
Error	25.252	140	.180		
Total	3013.960	143			
Corrected Total	33.905	142			

a. R Squared = .255 (Adjusted R Squared = .245)



## Build Public Trust

### Tests of Between-Subjects Effects

Dependent Variable: BPT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3.810 <sup>a</sup>	2	1.905	13.336	.000
Intercept	4.074	1	4.074	28.521	.000
PBT1	3.235	1	3.235	22.648	.000
RESPTYPE	6.577E-02	1	6.577E-02	.460	.499
Error	20.000	140	.143		
Total	3065.350	143			
Corrected Total	23.810	142			

a. R Squared = .160 (Adjusted R Squared = .148)

# APPENDIX H

## APPENDIX H

### Analysis of Covariance Testing: Hypothesis 5

#### Between-Subjects Factors

		N
RESPTYPE	P	73
	S	70

P = peers

S = subordinates

N = 73 for peers and 70 for subordinates

#### Overall Measurement

##### Tests of Between-Subjects Effects

Dependent Variable: ALL2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.035 <sup>a</sup>	2	3.517	28.424	.000
Intercept	4.235	1	4.235	34.224	.000
ALL1	5.523	1	5.523	44.632	.000
RESPTYPE	.445	1	.445	3.592	.060
Error	17.324	140	.124		
Total	2835.240	143			
Corrected Total	24.359	142			

a. R Squared = .289 (Adjusted R Squared = .279)

#### Steps to Empowerment

##### Tests of Between-Subjects Effects

Dependent Variable: STOE2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.694 <sup>a</sup>	2	3.847	26.988	.000
Intercept	6.390	1	6.390	44.825	.000
STOE1	6.229	1	6.229	43.699	.000
RESPTYPE	.469	1	.469	3.292	.072
Error	19.957	140	.143		
Total	2725.310	143			
Corrected Total	27.651	142			

a. R Squared = .278 (Adjusted R Squared = .268)

## Credibility

### Tests of Between-Subjects Effects

Dependent Variable: CRED2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11.564 <sup>a</sup>	2	5.782	37.609	.000
Intercept	3.941	1	3.941	25.637	.000
CRED1	9.766	1	9.766	63.526	.000
RESPTYPE	.312	1	.312	2.031	.156
Error	21.523	140	.154		
Total	2789.260	143			
Corrected Total	33.086	142			

a. R Squared = .349 (Adjusted R Squared = .340)

## Challenge and Support the Organization

### Tests of Between-Subjects Effects

Dependent Variable: CSO2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.955 <sup>a</sup>	2	2.478	23.016	.000
Intercept	3.821	1	3.821	35.493	.000
CSO1	4.085	1	4.085	37.949	.000
RESPTYPE	.239	1	.239	2.218	.139
Error	15.071	140	.108		
Total	3022.950	143			
Corrected Total	20.026	142			

a. R Squared = .247 (Adjusted R Squared = .237)

## Build Teams

### Tests of Between-Subjects Effects

Dependent Variable: BT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.430 <sup>a</sup>	2	3.215	17.403	.000
Intercept	10.320	1	10.320	55.862	.000
BT1	4.688	1	4.688	25.376	.000
RESPTYPE	.634	1	.634	3.433	.066
Error	25.863	140	.185		
Total	2809.580	143			
Corrected Total	32.293	142			

a. R Squared = .199 (Adjusted R Squared = .188)

## Develop Individuals

### Tests of Between-Subjects Effects

Dependent Variable: DI2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3.984 <sup>a</sup>	2	1.992	12.174	.000
Intercept	8.379	1	8.379	51.205	.000
DI1	2.761	1	2.761	16.874	.000
RESPTYPE	.627	1	.627	3.834	.052
Error	22.909	140	.164		
Total	2786.580	143			
Corrected Total	26.893	142			

a. R Squared = .148 (Adjusted R Squared = .136)

## Quality and Task

### Tests of Between-Subjects Effects

Dependent Variable: QT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.644 <sup>a</sup>	2	3.822	29.070	.000
Intercept	4.261	1	4.261	32.411	.000
QT1	6.171	1	6.171	46.938	.000
RESPTYPE	.511	1	.511	3.889	.051
Error	18.407	140	.131		
Total	2875.190	143			
Corrected Total	26.051	142			

a. R Squared = .293 (Adjusted R Squared = .283)

## Customer/Client

### Tests of Between-Subjects Effects

Dependent Variable: CC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.900 <sup>a</sup>	2	1.450	9.551	.000
Intercept	11.057	1	11.057	72.833	.000
CC1	2.380	1	2.380	15.675	.000
RESPTYPE	.276	1	.276	1.817	.180
Error	21.254	140	.152		
Total	2825.290	143			
Corrected Total	24.154	142			

a. R Squared = .120 (Adjusted R Squared = .107)

## Vision

### Tests of Between-Subjects Effects

Dependent Variable: V2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.511 <sup>a</sup>	2	3.756	22.775	.000
Intercept	7.972	1	7.972	48.346	.000
V1	5.754	1	5.754	34.893	.000
RESPTYPE	.750	1	.750	4.546	.035
Error	23.086	140	.165		
Total	2736.950	143			
Corrected Total	30.597	142			

a. R Squared = .245 (Adjusted R Squared = .235)

## Foster Creativity

### Tests of Between-Subjects Effects

Dependent Variable: FC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.549 <sup>a</sup>	2	4.774	28.097	.000
Intercept	6.249	1	6.249	36.772	.000
FC1	8.420	1	8.420	49.553	.000
RESPTYPE	.695	1	.695	4.088	.045
Error	23.790	140	.170		
Total	2759.740	143			
Corrected Total	33.339	142			

a. R Squared = .286 (Adjusted R Squared = .276)

## Manage Diversity

### Tests of Between-Subjects Effects

Dependent Variable: MD2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.733 <sup>a</sup>	2	3.866	25.253	.000
Intercept	5.294	1	5.294	34.576	.000
MD1	5.822	1	5.822	38.027	.000
RESPTYPE	.432	1	.432	2.819	.095
Error	21.435	140	.153		
Total	3000.100	143			
Corrected Total	29.168	142			

a. R Squared = .265 (Adjusted R Squared = .255)

## Build Public Trust

### Tests of Between-Subjects Effects

Dependent Variable: BPT2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.291 <sup>a</sup>	2	2.645	29.951	.000
Intercept	1.911	1	1.911	21.639	.000
PBT1	3.750	1	3.750	42.457	.000
RESPTYPE	.317	1	.317	3.585	.060
Error	12.365	140	8.832E-02		
Total	3113.860	143			
Corrected Total	17.656	142			

a. R Squared = .300 (Adjusted R Squared = .290)

## Vitae

David Sandt Corderman received his undergraduate degree from Lehigh University and masters' degrees from Virginia Commonwealth University and Virginia Polytechnic Institute and State University. He spent four years as an officer in the United States Marine Corps and then became an FBI Agent in 1983. Dr. Corderman currently serves as an instructor at the FBI Academy, where he teaches a course in Human Behavior in Organizations as part of the FBI's National Academy Program. He is also an adjunct professor for the University of Virginia and one of the managers of the National Executive Institute, an executive training program for chiefs of police with more than 500 sworn officers.