Introductory Total Quality Management Training

and

Supervisor and Subordinate Ratings of Supervisor Task Performance:

A University Administration Setting

by

Mary Minarovich-Cheniae

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

in

Industrial and Systems Engineering

1 June 95

Blacksburg, VA
Introductory Total Quality Management Training

and

Supervisor and Subordinate Ratings of Supervisor Task Performance:

A University Administration Setting

by

Mary T. Minarovich-Cheniae
H. A. Kurstedt, Jr., Chairman
Industrial and Systems Engineering

(ABSTRACT)

The purpose of this study was to determine what effects introductory Total Quality Management (TQM) training has on the differences in paired supervisor and subordinate ratings of supervisor task performance. I hypothesized the differences in ratings would decrease after the training.

This research was a quasi-experimental field study. My unit of analysis was established supervisor and subordinate pairs. Two established American university administrative offices provided the experimental (10 pairs) and control (6 pairs) groups. A Solomon Four Group Design was used; subjects were randomly assigned to pre-test, all other assignment was voluntary. There was no subject duplication. Each supervisor had a single subordinate and each subordinate had a single supervisor.

The treatment was a series of Deming Library video tapes followed by open discussion over the course of three months. Experimental group participation in the treatment was mandatory but response to the survey was voluntary.

The measurement instrument was broad: 118 task-related questions were clustered into 21 subscales based on previous work by the U.S. Merit Systems Protection Board. A seven-point scale was used.

A separate random sample of people associated with the university identified instrument subscales of greatest importance to the supervisor-subordinate relationship. The subscales were leadership, communication and supervision.

Primary analysis was a series of independent t-tests, comparing differences between paired differences. As anticipated due to the small sample size, few significant differences were observed. At post-test, a comparison of the supervisor and subordinate views differed. Experimental group subordinates tended to have more favorable views of their supervisors' performance while control group subordinates ratings tended to be less positive than the supervisor's views. The experimental group seems to have weathered the budget cuts with a more positive outlook.
Acknowledgment

I owe thanks to many people who have helped with my graduate school experiment. Without the assistance of the people mentioned below, I'm certain I could not have had as fulfilling an experience. Dr. Robert Dryden consistently located the funds that enabled me to remain in school for as long as I have. Mrs. Lovedia Cole remedied innumerable scheduling glitches, found me a quiet place to work, and generally smoothed my path. Dean Pamela Kurstedt provided the opportunity and setting for my research, shared her wisdom on prioritizing children and graduate school, and showered me with encouragement during the early stages of this effort. I thank you all for providing the essentials.

My committee has been quietly supportive. Dr. Harold Kurstedt agreed to chair my committee despite his overflowing calendar. Dr. Elizabeth Koball is a treasure, patiently listening as I traced a circuitous path to conclusions she knew from the start. Dr. Paul Torgersen was my willing advisor when I joined the department, after my committee was formed, and even after he became president. Dr. James R. Lang stepped in to help me when I needed help. I appreciate all the time and energy you shared with me.

My husband has taken care of so many of the small details for me: installing software, fixing unresponsive printers, converting my sketches into professional diagrams, delivering notes, preparing meals, making copies, driving great distances and learning the methodology of my research in order to help me with the programming. My work would have been immeasurably more difficult without his assistance. I don't think you'll ever know how much Clairisse and I thank you.
# TABLE OF CONTENTS

Chapter 1  INTRODUCTION  

| Sub-problems. | 1 |
| Outputs. | 1 |
| Relevance of the Study | 1 |
| Quality in the Workplace. | 1 |
| TQM on Campus. | 2 |
| This Study. | 2 |
| Who will Use this Study | 2 |
| My Study | 3 |
| Purpose. | 3 |
| Objective. | 3 |
| Research Question | 4 |
| Operational Research Question | 4 |
| Scope of the Study | 4 |
| Delimitations. | 4 |
| Premises. | 5 |
| Assumptions. | 5 |
| Operational Definitions. | 6 |
| Conceptual Model | 7 |
| Research Hypotheses | 8 |

Chapter 2  LITERATURE REVIEW  

| Total Quality Management | 10 |
| Is TQM New? | 10 |
| What Makes Deming Special? | 13 |
| American Universities | 14 |
| Is There Something Unique about a University Organization? | 14 |
| Don't American Universities Already Have High Quality Standards? | 14 |
| TQM and Universities | 14 |
| Which Academic Institutions Are Exploring TQM? | 15 |
| Are Deming's Ideas Appropriate for a University Administrative Office? | 15 |
| The Supervisor-Subordinate Relationship | 16 |
| What Is the "Relationship"? | 17 |
| Is the Importance of the Supervisor Exaggerated? | 17 |
| What about Communication? | 17 |
| Why Do We Care about Employee Satisfaction? | 18 |
| What Do Subordinates Expect of Their Supervisors? | 19 |
Perception
  How Does Perception Work? 20
  Are Self Evaluations Accurate? 21
  Are Evaluations by Subordinates Accurate? 21
  How Do Evaluations by Self and Subordinate Compare? 22
  Is There Value in Having Subordinates Rate Their Supervisors? 24
Training
  Learning 25
  Training Evaluation 27
My Conceptual Model 29
Chapter 3  METHODOLOGY 31
Research Design 31
Experimental Design 32
  A Solomon Four Group Design. 32
  My Design. 33
Population 34
Measurement Instrument 35
  My Measurement Instrument. 35
  My Modifications. 35
  The Questions. 36
  The Audience. 36
  Normative Data. 36
  Reliability Data. 37
Pilot Study 37
Subject Confidentiality 38
  Collecting the Data. 38
  Coding. 38
  Within the Thesis. 38
Treatment 39
Data Collection 39
  The Survey. 39
  The Procedure. 39
Interview with the Experimental Group Training Facilitator 40
Analysis Methods 40
  Focus of Analysis. 40
  Response Analysis. 40
  Threats to Validity. 41
NOTES 42
Chapter 4  RESULTS

Focus of Analysis

My Interview with the Experimental Group Training Facilitator. 44
The Secondary Survey. 46

Primary Survey

Responses. 47
Measurement Instrument Internal Consistency. 47

Research Hypotheses Testing

There Will Be a Significant Difference between Experimental and Control Group Differences for the Post-test. 48
There Will Be no Significant Difference between the Experimental and Control Group Differences for the Pre-test. 49
The Instrument Will Have an Effect on the Subjects. 50
Differences between the Supervisor and Paired Subordinate Responses Are Affected by the Treatment. 51
Experimental Group Subjects Who Provide Responses for the First Time Will Differ between Pre- and Post-tests. 52

NOTES 54

Chapter 5  EXPLORATORY ANALYSIS

Comprehensive Analysis

Supervisors Don't Always Rate Themselves More Favorably Than Their Subordinates Rate Them. 55
Significant Differences. 65
Fourteen Most-Important Tasks. 68
General Impressions. 69

Some Observations

The Doesn't Do This Task Response. 70
The I Can't Judge Response. 70
The Can Do This Task Exceptionally Well Response. 70

Chapter 6  DISCUSSION

Interpretation of My Results

Instrument Reliability. 71
Hypothesis Testing. 73
Pairs Completing Both the Pre- and Post-tests. 75
Supervisors Don't Always Rate Themselves More Favorably Than Their Subordinates Rate Them. 76

Improvements I Would Make

Pilot Study. 77
Evaluation Scale. 77
Procedures. 78
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>My Conceptual Model.</td>
<td>7</td>
</tr>
<tr>
<td>Figure 2</td>
<td>My Research Hypotheses.</td>
<td>9</td>
</tr>
<tr>
<td>Figure 3</td>
<td>My Research Design Model.</td>
<td>32</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Mean Pre-test Subscale Scores for Experimental Group Subjects.</td>
<td>56</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Mean Pre-test Subscale Scores for Control Group Subjects.</td>
<td>56</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Mean Post-test Subscale Scores for Experimental Group Pre-test and Post-test Subjects.</td>
<td>57</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Mean Post-test Subscale Scores for Control Group Pre-test and Post-test Subjects.</td>
<td>57</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Mean Post-test Subscale Scores for Experimental Group Post-test Only Subjects.</td>
<td>58</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Mean Post-test Subscale Scores for Control Group Post-test Only Subjects.</td>
<td>58</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Median Pre-test Subscale Scores for Experimental Group Subjects.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Median Pre-test Subscale Scores for Control Group Subjects.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Median Post-test Subscale Scores for Experimental Group Pre-test and Post-test Subjects.</td>
<td>61</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Median Post-test Subscale Scores for Control Group Pre-test and Post-test Subjects.</td>
<td>61</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Median Post-test Subscale Scores for Experimental Group Post-test Only Subjects.</td>
<td>62</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Median Post-test Subscale Scores for Control Group Post-test Only Subjects.</td>
<td>62</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Absolute Differences between Paired Supervisor and Subordinate Median Scores for the Fourteen Most-Important Tasks.</td>
<td>68</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Absolute Differences between Paired Supervisor and Subordinate Median Scores for General-Impression Items.</td>
<td>69</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1  Solomon Four Group Design.  33
Table 2  Design of My Study.  33
Table 3  Secondary Survey Results.  46
Table 4  Descriptive Statistics of Responses.  47
Table 5  Cronbach’s Alpha for Each Subscale.  47
Table 6  Design of the Post-test T-test.  48
Table 7  Descriptive Statistics for the Post-test T-test Responses.  48
Table 8  Post-test T-test Results.  49
Table 9  Design of the Pre-test T-test.  49
Table 10 Descriptive Statistics for the Pre-test T-test Responses.  49
Table 11 Pre-test T-test Results.  49
Table 12 Design of the Pre-test and Post-test versus Post-test Only T-test.  50
Table 13 Descriptive Statistics for the Pre-test and Post-test versus Post-test Only T-test Responses.  50
Table 14 Pre-test and Post-test versus Post-test Only Post-test T-test Results.  51
Table 15 Design of the Pre-test Post-test Differences T-test.  51
Table 16 Descriptive Statistics for the Pre-test Post-test Differences T-test Responses.  52
Table 17 Pre-test Post-test Differences T-test Results.  52
Table 18 Design of the First Time Experimental Group T-test.  53
Table 19 Descriptive Statistics for the First Time Experimental Group T-test Responses.  53
Table 20 First Time Experimental Group Participants T-test Results.  53
Table 21 Descriptive Statistics of Responses.  59
Table 22 Pre-test Subscale Ranking Results.  63
Table 23 Descriptive Statistics of Pre-test Paired Difference Responses with Significant Differences at $\alpha = 0.05$.  63
Table 24 Post-test Subscale Ranking Results.  64
Table 25 Descriptive Statistics of Post-test Paired Response Differences with Significant Differences at $\alpha = 0.05$.  64
Table 26 Subscale T-test Results.  66
Table 27 Dependent T-test Results.  67
Chapter 1
INTRODUCTION

Chapter 1 explains the background of my study. The study was conducted in parallel with a training and development program, not in a laboratory setting. My scope is carefully defined. The conceptual model is briefly described. All my hypotheses are provided.

Problem Statement
This study measures how introductory Total Quality Management (TQM) training affects supervisor and subordinate perceptions of supervisor task performance for established supervisor-subordinate relationships in an administrative sub-organization of an American land-grant university. I'm interested in measured differences within the pairs of supervisors and subordinates.

Sub-problems
1.) Before introductory TQM training, do supervisors and subordinates have similar perceptions of supervisor task performance?
2.) Does introductory TQM training affect supervisor and subordinate perceptions of supervisor task performance?
3.) Is this study repeatable?
4.) Does this study provide guidance for future studies of introductory TQM training?

Outputs.
1.) Descriptive statistics of responses to the pre-test and post-test.
2.) Results of the effect of training on differences observed in the pre-test and post-test.
3.) Improvements I would make if I were to do this study again.
4.) Recommendations for managers considering TQM training and training evaluation.

Relevance of the Study
Quality in the Workplace. Hardly a newspaper or trade journal can be read without finding an article on quality. Global competition has been one of the drivers of this pursuit of quality. This international perspective is reflected in recent curriculum content changes for various programs of study at some American universities. The topics of quality and TQM arise in lecture halls and conference rooms for many academic departments.
There are multiple approaches toward quality. TQM is one of them. Dr. W. Edwards Deming was a widely recognized teacher of TQM theory at the time this study was conducted. After gaining a foothold in industry applications, TQM concepts are being applied to administrative functions in government and academia.

*TQM on Campus.* The impetus for quality in a university setting isn’t only ingraining a quality awareness in the students, enabling them to apply TQM in their future careers, but also instilling a quality awareness among the faculty and staff. Tighter fiscal constraints and new student demographic trends (Fife & Barrette, 1986; Marchese, 1991; Seymour, 1993) provide additional momentum for academic leaders to adapt and modify these teachings within their own organizations (Lewis & Smith, 1994).

Industry is prodding academia. In 1991, the leaders of American Express Company, IBM Corporation, The Proctor and Gamble Company, Ford Motor Company, Motorola, and Xerox Corporation collaborated on a public statement regarding TQM. "We believe business and academia have a shared responsibility to learn, to teach, and to practice total quality management. If the United States expects to improve its global competitive performance, business and academic leaders must close ranks behind an agenda that stresses the importance and value of TQM" (Robinson et al, 1991, p. 94). Academic institutions are rising to meet this challenge. TQM initiatives are underway at universities throughout the country.

*This Study.* My opportunity to study the effects of introductory TQM training on subjects in their workplace is the result of an American university administrative office’s initiative to try to instill a TQM philosophy. Learning to do more with less, concern for employee morale, and a desire to work smarter prompted this sub-organization to invest in Deming’s TQM training. After several managers attended a Deming seminar, a decision was made to purchase video tapes from the Deming Library. I was invited to conduct a scientific study in conjunction with the sub-organization’s evaluation of the training.

Who Will Use this Study

I think this study will be helpful to more than university administrators. There is a dearth of rigorous evaluation of the effects of TQM training despite the fact that many organizations are investing heavily in such training. The university population is unique, but perhaps similarities
among administrative functions in different organizations, from a practitioner's perspective, will cause this study to help an assortment of organizations.

The training was directed at issues common to many organizations. Example issues are restructuring office functions and improving work. If administrators, project managers and other white collar workers are looking for tools to tackle similar issues, they will benefit from this study. Seeing what happened in one office can be a useful guide for another organization. When managers understand the relationship between introductory TQM training and supervisor-subordinate relationships, they will have realistic expectations for their training ventures. TQM training investment decisions will be rational.

University administrators will benefit most from this study because potential pitfalls unique to the population of American university administrators, useful for their own introductory TQM training and evaluation efforts, are identified as part of this study.

My Study

Purpose. Many organizations are imposing training programs on their employees without identifying goals for that training. Because introductory TQM training has been conducted in many successful organizations, managers looking for improvements in their own domains prescribe such training without analyzing whether such training will affect the areas they are trying to improve. This research will determine whether there is a relationship between introductory TQM training and supervisor and subordinate perspectives of supervisor task performance. I compare differences in ratings between matched pairs of real-world supervisors and subordinates before and after training and with and without training. I'm doing this study because I want to know whether introductory TQM training more closely aligns supervisor and subordinate perceptions. Since I assume greater similarity of supervisor and subordinate perceptions are beneficial to an organization, I'm hoping to see increased alignment after the training is completed.

Objective. My objective is to determine whether differences between supervisor and subordinate perceptions of supervisor task performance in the course of performing normal duties are affected by introductory TQM training. These perceptions are measured with the ratings assigned by the supervisor and subordinate.
Research Question
How does introductory TQM training for supervisors and their subordinates change the differences between a supervisor's perspective of the supervisor's task performance and the subordinate's perspective of that supervisor's task performance as measured pre- and post-training?

Operational Research Question
If an American university administrative sub-organization invests time and energy introducing all its members to the concepts and tools of TQM, will supervisors and subordinates become more similar in their perceptions of supervisor task performance?

This operational research question can be divided into sub-questions:
1.) How do the experimental and control groups differ after the treatment?
2.) How do the experimental and control groups differ prior to the treatment?
3.) Does the measurement instrument have an effect on paired response differences?
4.) Are differences between the supervisor and paired subordinate affected by the treatment?
5.) Can the effects of the treatment be observed within the experimental group?

Scope of the Study

Delimitations.
1.) Due to confidentiality concerns, this experiment doesn't explore possible differences in relationships due to race, gender, age, education or any demographic criteria. Because of the small sample sizes and the ease in identifying individuals simply by scanning such data, no demographic information was solicited from the subjects.
2.) This experiment focuses on supervisor-subordinate pairs. I don't consider either the organizational structure within the two sub-organizations or organizational effectiveness.
3.) I'm concerned only with the supervisor's and subordinate's perceptions of the supervisor. I'm not concerned with either reciprocal views or how supervisors think they are viewed. The instructions at the top of my instrument state this.
4.) This experiment isn't concerned with the merits or shortcomings of the training program used as treatment for the experimental group of subjects. I'm only interested in the effects of the training on the perceptions of the subjects. The format, content, and timing are controlled by the initiative of the sub-organization.
5.) The instrument asks for perceptions in the "now". The analysis is concerned with the differences in the supervisor's and subordinate's perceptions of the supervisors. The pairs of responses for any given question are compared for the same time.

6.) Natural groups, not random assignment, determined the composition of the experimental and control groups.

Premises.

1.) For this experiment, both sub-organizations are part of the same parent organization. External forces are ignored: I assume university-wide policies, directives and global issues impact the two offices simultaneously.

2.) The perceptions of paired supervisors and subordinates are reflected in their ratings of the supervisor.

3.) This research is a quasi-experiment.

Assumptions.


2.) The training was adequate. Dr. Deming conducted seminars expounding his theories until his death. The Deming Library is a series of video tapes, endorsed by Dr. Deming, developed to teach TQM concepts to a wide range of audiences. Over the course of three months, the experimental group watched ten training tapes, each followed by an open discussion, and participated in a workshop. The basic theories and tools of TQM were introduced and practiced. The training facilitator and the top manager in the experimental group considered this sufficient background to achieve their goals. While these ten video tapes aren't the entire library, each is a self-contained course and all were individually selected to meet the needs of the experimental sub-organization. The top manager and training facilitator of the experimental group considered these tapes sufficient introductory TQM training for their purposes based on their knowledge of the people in the sub-organization. Criteria used to select the video tapes included teamwork, process flowcharting, and a systems approach towards problem solving.

3.) A closer alignment of supervisors and their subordinates is beneficial to the workplaces is an intuitive assumption; I couldn't find any literature supporting or arguing against this belief. It seems logical to suppose that for supervisors and subordinates to view things in the same way, they
must be able to communicate and cooperate - there must be a common ground for them to see things in the same way. I assume the organization benefits from this. Graen & Schieman (1978) found that agreement between supervisors and subordinates on the meaning of mutually experienced events is a positive function of the quality of their relationship.

4.) Administrative Government employees are the same. Distinctions between Federal government employees and land grant university employees were ignored. Both types of employees are government workers; both are knowledge workers. The originators of the instrument, comparing their studies with previous studies of management responsibilities, found "the substance of the findings was very similar" (Flanders and Utterback, 1985, p. 404).

Operational Definitions. These terms are used throughout my thesis:

Sub-organization: relatively autonomous college administrative office. Major responsibilities include the interpretation and implementation of university-wide policies. Within university guidelines, internal organizational structures and procedures differ between sub-organizations.

Subordinate: individuals supporting people directly responsible for defined areas in the sub-organization domain. Subordinates provide clerical assistance, accounting functions, or program support. All have well-established relationships with their supervisor. All are full-time employees.

Supervisor: individuals appearing on the organizational charts for the sub-organizations. Each supervisor has primary responsibility for some function within the office. All supervisors have well-established relationships with their subordinates. All are full-time employees.

Total Quality Management: "TQM is a comprehensive philosophy of living and working in organizations that emphasizes the relentless pursuit of continuous improvement" (Chaffee & Sherr, 1992, v). The concepts comprising TQM are not new and exist under many guises, even the basic tools have a long history, but the packaging of the tools and concepts has gained recent attention in government, business, and academia as organizations strive to remain competitive. "If you don't empower people for quality, the organization becomes a collection of underachievers" (Seymour, 1993, p. 101).

Survey: a packet containing completion instructions, a copy of the appropriate instrument, and a coded opscan form. Instrument wording differed slightly between experimental and control groups
and supervisors and subordinates although the intent of the questions remained consistent. Opscan forms contained supervisor or subordinate codes and pairing codes when first assigned.

*Training:* the treatment for this study.

*Rating:* instrument responses marked on the opscan form. How I measure perceptions of supervisors and subordinates.

Conceptual Model
Supervisor task performance has been conducted for an undetermined amount of time. Both supervisors and subordinates have opinions regarding the supervisor's task performance. If these perceptions can be measured, the differences between them can also be measured. Both supervisors and subordinates bring their perceptions to the introductory TQM training. The training will have some impact on their perceptions after the training is completed. These perceptions are also affected by the supervisor's task performance. When these perceptions are measured again, the differences will also be measured. Figure 1 shows the relationship between the supervisor's task performance and supervisor's and subordinate's perceptions of that supervisor's task performance and introductory TQM training and supervisors and subordinates perceptions. My study stops short of examining the extent of change in supervisor task performance. Like Watson, I believe learning is demonstrated in expressed attitudes. In order to be complete, my conceptual model steps beyond the scope of my study.

![Diagram of conceptual model](image)

**Figure 1.** My Conceptual Model.
Research Hypotheses

When I began my study, I had an expectation based on the literature review. That I expected supervisors to evaluate themselves more favorably than their subordinates evaluated them has been apparent throughout the evolution of my thesis. While this idea as a testable hypothesis is outside the scope of my study, it has always been my intention to verify whether or not my data support this expectation. My focus is on the differences between paired supervisor and subordinate perceptions as measured by the ratings assigned in the evaluations.

The concepts behind my hypotheses can be seen in Figure 2. Within the university there are two sub-organizations, the experimental group (the upper part of the diagram) and the control group (the lower part of the diagram). Both groups participate in a pre-test at about the same time. The experimental group receives introductory TQM training. Later, both groups participate in a post-test at about the same time.

My anticipation of the relationships among the subjects' ratings are portrayed in the diagram. The interpretation of the diagram is general — the relationship between supervisor and subordinate ratings is my focus. I'm trying to show when I expect to see the difference between supervisor and subordinate ratings decrease or remain the same. I don't attempt to predict exact ratings or differences. To draw attention to the general trend I expect to observe, I use the entire seven point scale. I expect supervisors will always evaluate themselves more favorably than their paired subordinates evaluate them. In the pre-test, I don't expect to see a significant difference between the experimental and control groups. In the post-test, I don't expect the two groups to have similar rating patterns. I expect the difference between supervisors and subordinates to decrease for the experimental group and to remain unchanged for the control group. My dependent variable is the difference between paired supervisor and subordinate ratings.

My operational research sub-questions can be translated into testable hypotheses:

1.) There will be a significant difference between the experimental and control group differences for the post-test.

2.) There will be no significant difference between the experimental group and control group differences for the pre-test.

3.) The instrument will have an effect on subjects.

4.) Differences between the supervisor and paired subordinate are affected by the treatment.
5.) Experimental group subjects who provide responses for the first time will differ between the pre- and post-tests.

Figure 2. My Research Hypotheses.
Chapter 2
LITERATURE REVIEW

Chapter 2 describes the literature that helped me with my study, including the formation of my hypotheses. In addition to traditional academic references sources, I examined periodicals directed at practitioners. Since my setting is a real-world environment, I looked at the information managers have ready access to. I've included only articles providing substantive insight, guidance or theory, or demonstrating some degree of scientific methodology.

Basically, my study is the evaluation of the implementation of a training program with a few added twists. I use two groups of subjects: experimental and control. Before the treatment, approximately half the subjects in each group provide responses to a survey. After the treatment, all the subjects provide responses to the same items as the original survey. For the experimental group, the treatment is introductory TQM training. For the control group, there is no intervention. The measurement instrument addresses supervisor task performance. Supervisors in both groups are asked to evaluate their own task performance. Subordinates in both groups are asked to evaluate their supervisor's task performance. The basis of all these evaluations is the perceptions of the rater. For this study, I don't observe and note the supervisor's performance nor do I examine the differences between the two perceptions of that performance before and after the introductory TQM training.

Contrary to most research papers, I'll save the literature directly related to my conceptual model for the end of this chapter. I want to set the stage carefully.

Total Quality Management
Although TQM theory and practice aren't directly related to this study, they are the foundation my study is built upon. Experimental group management considered TQM the appropriate approach for their office, committing money and time. Ten tapes from the Deming Library were purchased. Criteria used to select the tapes included teamwork, process flowcharting, and a systems approach towards problem solving. Attendance by all members of the office was mandatory.

Is TQM New? TQM is one packaged approach for instilling a quality mentality. W. Edwards Deming, respected TQM lecturer, is often recognized as the originator of TQM. While Deming didn't like the term TQM, he's usually credited with the initiation of the quality movement in the
United States. This movement is labeled TQM. In reality, some of the themes inherent in TQM have long histories. I see the similarities among writers with a people-orientation toward quality more readily than I can identify the differences. My attitude may be the result of living through several quality program fads while I was working as an electronics engineer in large bureaucracies. Deming’s theories further the beliefs of some noted thinkers and are consistent with what I observed in the workplace. I want to explain why I appreciate Dr. Deming’s approach toward a quality mentality by describing the lineage of some of his ideas. Since the training used in my research was based on the Deming tapes, the effect of the training on supervisors and their subordinates was influenced by Deming’s approach.

Deming reiterates the value of people again and again in his teaching. This view isn’t unique. There are enough twentieth century authors with a humanistic bent for me to illustrate the sound foundation upon which TQM rests. The key to a better workplace has been elusive, but common themes keep reappearing. I’ll limit my discussion to organizational theories, intentionally omitting any discussion of tools or statistical theory since my research is based on human perceptions. Of the many published thinkers, I’ve chosen three: Mary Parker Follett because she is the earliest author I can identify who combines an empathy for the individual with a perceptive understanding of the motivations and intrigue within an organization; Douglas McGregor because his work is widely recognized as an enormous contribution to so many applied fields; and Peter F. Drucker because he’s a well-established, widely read observer and philosopher of American organizations. The dilemmas these authors address are the same as the issues the experimental group wanted to tackle with introductory TQM training.

Follett (1940) emphasizes and reemphasizes the need for co-ordination; the integration of different views makes an organization stronger. Employees must understand the functional unity and interdependence of all activities to have an intelligent opinion of their own concerns within the organization. Managers are teachers. Management should mean sensible working arrangements, not control of the employees. "Authority and responsibility should not depend on hierarchy of rank (Follett, 1940, p. 173)," but should go with function; subordinates are sometimes managing. This attitude was reflected in the Deming library tapes used for the treatment in this study. These tapes were selected to teach an appreciation of how the work of one person affects the work of others.

McGregor is perhaps most famous for Theory Y Management: "to encourage integration, to create a situation in which a subordinate can achieve his own goals best by directing his efforts towards
the objectives of the enterprise" (1960, p. 61). He associated with a number of similarly minded people and his views are reflected widely in organizational and management literature. He cautions that the supervisor will not help the subordinate or the organization "if he attempts to keep direction and control in his own hands; he will only hamper their growth and encourage them to develop countermeasures against him" (1960, p. 152). Growth and development of the subordinate, a responsibility of the organization, includes opportunities to express ideas and suggestions on matters affecting the subordinate before the decisions are made, not only helping the subordinate to learn but providing genuine satisfaction (1944). From another angle, Deming urges the participation of every person in an organization. Deming persuades management by arguing the failure to use the abilities of every worker is a waste of resources.

The ideas of Peter Drucker, a prolific writer for several decades, have evolved over time. His position has become more finely tuned. For my research I'm primarily interested in his thoughts on the "knowledge worker" since that is an apt way to think of people working in administrative capacities. He reminds us that "specialized knowledge has no results unless it focuses on the needs and goals of the entire organization" (1989, p. 96). "The more knowledge-based an institution becomes, the more it depends on the willingness of individuals to take responsibility for contribution to the whole, for understanding the objectives, the values, the performance of the whole, and for making themselves understood by other professionals, the other knowledge people, in the organization" (1989, p. 97). He discusses the "satisfied customer" as the product of such an organization. A knowledge worker may have a supervisor and may have subordinates. These ideas are reflected in Deming's opinions. He wants people to understand what is performed in other positions and how staff positions should be optimized for the organization and not for the sub-organization.

The ideas of Follett, McGregor and Drucker are universal: each of these topics is mentioned in my interview with the intervention facilitator, yet the paragraphs above were written months before the interview was conducted. It was these issues the experimental group hoped to tackle with introductory TQM training.

All these ideas are consistent with current TQM literature. "The aim of leadership should be to improve the performance of man and machine, to improve quality, to increase output, and simultaneously to bring pride of workmanship to people" (Deming, 1986, p. 248). "TQM highlights the interdependence of work and the importance of involving all affected parties in
making improvements" (Chaffee & Sherr, 1992, p. 68). Marchese points out "the prime 'customer' of most administrative offices turns out to be . . . other administrative offices" (1991, p. 6). "If you don't empower people for quality, you will become inundated with bureaucrats and policy manuals" (Seymour, 1993, p. 100).

I don't believe an organization has to choose one approach toward a quality mindset over another. The terminologies and steps for attaining quality vary, but there are so many similarities between the end products that when the organization is living and breathing quality, I think the theory is transparent. Stieber (1991) discerned no differences in organizational reactions to a number of quality approaches applied to different real-world organizations. Lewis & Smith (1994) have a framework for reconciling the various approaches toward quality; they consider an organization to be composed of three systems: social, technical and management. In their view quality also has three perspectives: Deming presents the social perspective; Juran presents the technical perspective; and Ishikawa presents the management perspective.

It's important to choose a teacher able to communicate ideas with the audience. The Deming library was created to help many audiences. I assume the content and teaching methods were appropriate for the audience.

*What Makes Deming Special?* Dr. Deming is credited with revolutionizing post-war Japanese industry. His achievements in the transformation are recognized with the annual Union of Japanese Science and Engineering quality improvement award, the Deming Prize. His theories, tools and techniques have a proven track record.

In *Out of the Crisis* (1986) Dr. Deming proposes his ideas on transforming American management. His statistical background is apparent in his discourses on variability, but isn't pertinent to my study. The video tapes of the Deming Library delve into his ideas, interspersing tutoring with example applications and interviews with Dr. Deming; these tapes are endorsed by Dr. Deming as teaching aids for many audiences. The most famous feature of his theories on quality may be his "14 Points for Management". He stresses the need for constancy of purpose: everybody knowing what to do and everybody doing their best. He repeats the need to drive out fear so people can be effective. And he addresses the barriers and handicaps that rob people of pride of workmanship. The word "teams" frequently appears in his work.
American Universities

A university campus is not a microcosm of the world. The environment is more than just the place where students attend classes, it's an incubator for ideas and a laboratory for experimentation. Students are exposed to new situations, learning in unexpected places. The smallest experience, frustrating or satisfying, could be a permanent memory for a student. I think this sense of readiness and acceptance is contagious; the feeling of energy I sense on campus is far different from what I experienced in the working world.

*Is There Something Unique about a University Organization?* The organizational structure of colleges and universities presents some difficulty in developing a focused institution-wide effort. Not only do the schools and departments operate fairly autonomously, a dual structure exists. There are separate but meshed hierarchies for administrative and academic functions, resulting in intensive divisionalization and fragmented leadership. Rather than accept this difficulty as justification for the status quo, Lewis and Smith (1994) consider it even more reason to pursue total quality management on campus. "If your colleagues and other members of your institution don't use virtually the same words to describe your objectiveness, your individuality, you have a problem that necessarily affects the quality of your operation." suggests Seymour (1993, p. 60).

*Don't American Universities Already Have High Quality Standards?* Players in the academic arena have convinced themselves they already have a quality perspective. There are criteria for admitting students and criteria for grades. Research methodology concepts are established and peer review is the norm prior to publication. But Lewis and Smith (1994) question the assumption that higher education is quality driven. There is a push for accountability and meeting expectations. In their view, "total quality is a set of philosophies by which management systems can direct the efficient achievement of the objectives of the organization to ensure customer satisfaction and maximize stakeholder value" (Lewis & Smith, 1994, p. 29).

TQM and Universities

Not all institutions of higher learning are experimenting with quality. In varying degrees, quality is addressed in business and engineering programs; students are exposed to the basic concepts in many of their classes. Since I've returned to school, I've attended classes in both colleges at the graduate and undergraduate levels; the topic of quality cropped up in most of these courses.
Which Academic Institutions Are Exploring TQM? A cursory look at documented TQM experiments conducted on post-secondary campuses in America shows three major players: community colleges, business schools and engineering schools. Marchese observes "notably absent from TQM rosters are liberal arts colleges" (1991, p. 8).

Most of the TQM literature from academia is either case studies or how-to articles. Since experienced researchers conducted these initiatives, it's natural they would publish their observations, particularly since so many business and industry leaders are calling for TQM in education. Unfortunately, most of these publications were not scientifically documented, so little could be gleaned for my study although they may provide some assistance to future practitioners.

Industry is clamoring for TQM at the university level. "Our system of higher education is one of this country's most powerful competitive weapons. Working together, companies and institutions of higher education must accelerate the application of total quality management on our campuses if our education system and economy are to maintain and enhance their global position" (Robinson et al, 1991, p. 95).

Taylor (1990) presents a persuasive case for TQM in engineering education. Since it's more than a set of criteria to be met, the learning environment, which has significant impact on performance, will benefit from the right blend of academic and social circumstances. Students learn what they experience and observe. When students are living a TQM mentality, they will carry TQM through to their careers.

Are Deming's Ideas Appropriate for a University Administrative Office? Dr. Deming sees no distinction between manufacturing and service organizations. While he doesn't specifically say TQM is appropriate for university administrative offices, he uses comparable illustrations in his book and interviews. Many administrative functions are common among organizations.

Dr. Deming's ideas have generated a great deal of interest in managers in just about every kind of organization imaginable. A number of books linking "quality" and "education" have been published. Most seem to rehash Dr. Deming's words, often quoting him endlessly without contributing anything of substance for the reader. One notable exception was written by Chaffee and Sherr, Quality: Transforming Postsecondary Education; this intelligent book examines the
concepts of TQM and discusses the ramifications for educators and academic institutions. The essence of TQM is simplified to three ideas: defining quality, improving the organization's work performance and improving the organization's administrative performance. Deming's fourteen points are "especially well-suited to post secondary settings" (Chaffee & Sherr, 1992, p. 3). "Administration's job is to create an environment where people can work productively and effectively. Administrators must ensure adequate training and education. People cannot prevent error unless they know what constitutes error and how to avoid it" (Chaffee & Sherr, 1992, p. 67).

A few authors step beyond the mirroring of Dr. Deming's words. Seymour (1993) is quite comfortable delving into philosophy, making observations and drawing conclusions. His discussions aren't really devoted to higher education, they are appropriate for any organization. His insight and ease contributed to my understanding, but I never felt he has intimate knowledge of a campus. Marchese (1991) provides some insight and demonstrates time spent thinking about TQM on campuses. Chaffee & Sherr (1992) move beyond TQM theory and demonstrate personal understanding of the workings of an institution of higher education.

A TQM mentality is an asset to an institution of learning. "Viewed broadly, TQM is a call to quality and a mindset about improvement; it values data, teams, and process; for the many faculty and staff of an institution, it offers respect and a voice. As for its special tools and vocabularies, they come second and are for adaptive use" (Marchese, 1991, p. 6).

The Supervisor-Subordinate Relationship

My study is based on supervisor-subordinate pairs. I'm interested in differences between ratings assigned by the two members of each pair. I analyze these differences. All of my methodology and design is built upon supervisor-subordinate pairs.

The literature supports what I've observed in the workplace. The attitudes and actions of an individual supervisor have a great deal of impact on subordinates, whether positive or negative, and the organization feels the effects. Like a pebble tossed into a pond, the ripples spread beyond the original point of contact.

The supervisor-subordinate relationship is undoubtedly the most thoroughly explored area I delved into. While I found some studies confirming the results of previous studies, many are minutely
incremental verifications of theory, with only subtle conclusion distinctions. Two themes appear repeatedly in supervisor-subordinate discussions: communication and satisfaction.

*What Is the "Relationship"?* There is an interpenetration of responsibility, a collective responsibility among the members of an organization, with a circular nature to the interaction within the supervisor-subordinate relationship (Follett, 1940). Within this relationship, there are dimensions of trust, respect, loyalty, liking, intimacy, support, openness and honesty (Graen & Scandura, 1987). An organization is structured on the summation of these relationships. McGregor points out "every manager at every level is dependent upon those below him in the organization" (1960, p. 23).

The relationship isn’t one-sided: both people feel its impact. Herold (1977) documented the interrelationship between supervisor and subordinate: supervisor behaviors and attitudes are a function of subordinate performance while subordinate behaviors and attitudes are a function of supervisor behaviors. Supervisor behavior as perceived by subordinates is substantially related to the productivity of the work group (Lawshe & Nagle, 1953).

The quality of the supervisor-subordinate relationship is influenced by performance feedback, organizational goals and job rationale (Cooley, 1986). The nature of supervisor-subordinate "relationships in an organization, as perceived or evaluated by subordinates, affects the job performance of subordinates" (Bernard et al, 1961, p. 357).

*Is the Importance of the Supervisor Exaggerated?* "It has become a generally accepted fact that well-qualified first-line supervisors are a vital ingredient in the chemistry of successful organizations" (Merit Systems Protection Board, 1992, p. 1). McGregor considers the supervisor-subordinate relationship to be the most important relationship in industry (1944). These opinions are supported by empirical research. Herzberg et al (1957) reviewed fifteen studies of more than 28,000 employees and found that the second most frequently mentioned factor for satisfaction is the subordinate’s relationship with the supervisor.

*What about Communication?* The most important communication taking place in an organization is within the supervisor-subordinate relationship.
Subordinates who have good communication with their supervisors tend to be better informed and are more satisfied with what is happening in the organization (Lesniak, 1981). Supervisor information directly associated with the job is most likely to be accurately communicated without distortion (Level & Johnson, 1978).

Bernard et al (1961) believe the channel of communication between supervisors and subordinates is positively related to performance. Supervisor-subordinate communication is desirable in terms of organizational effectiveness (Cahn, 1986). Supervisor communication effectiveness impacts subordinate performance (Jenkins, 1977): feedback from supervisors is important to U.S. workers (Earley, 1986). Rippey (1980) found that communication within the supervisor-subordinate dyad not only impacts the organization, it's positively correlated with the performance evaluations the subordinates receive.

The freedom of communication within the supervisor-subordinate relationship is positively related to performance (Bernard et al, 1961). Communication within the relationship flows two ways: from the supervisor to the subordinate and from the subordinate to the supervisor. In turn, both are transmitters and both are receivers. Deming (1986) and Ryan and Oestreich (1991) consider fear of voicing beneficial contributions an all too common waste of an organization's assets, an opinion supported by several authors. Landen believes an organization "must have the full support and commitment of every member" (1989, p. 79). Chaffee and Sherr are even more compelling: "An effective organization requires the best and smartest efforts of everyone of its personnel. Anything less is waste, lost productivity" (1992, 64). Satisfaction with supervisor receptivity is a powerful predictor of subordinate job satisfaction (Wheeless et al, 1984).

Bernard et al (1961) found consistent evidence that open communication within the supervisor-subordinate relationship and a high degree of mutual understanding of viewpoints and problems, are positively correlated with a high level of performance. "Communication effectiveness is a key element in productivity and job satisfaction" (Rybczyk, 1987). Strong supervisory communication skills lead to employee satisfaction (Rybczyk, 1987; Plunkett, 1982).

Why Do We Care about Employee Satisfaction? There is a popular belief that when employees are involved in workplace decisions, improved quality, productivity, motivation and satisfaction result (Lander, 1989). Job satisfaction and job performance are positively correlated (Petty et al, 1984). While Dr. Deming doesn't directly address employee satisfaction, the issue is heavily
alluded to in his discussions of fear, pride in workmanship, self-improvement, teamwork, leadership and supervision, elimination of exhortations and quotas, and the new way of thinking (Deming, 1986).

Subordinates feel satisfaction when they know their opinions and ideas are welcome contributions (McGregor, 1944). The "Golden Rule", a term coined by Follett (1940), says that when employees feel like partners, their interests are the same as management’s interests. If a subordinate "is made aware of the significance of his or her work - its impact on the rest of the company - it'll actually inspire the person to build himself or herself up" (D’O’Brien, 1992, p. 9).

The supervisor-subordinate "relationship appears to be central to overall satisfaction and organizational effectiveness" (Cooley, 1986, p. 2). Subordinate attitude toward a job situation correlates with attitude toward interpersonal relations with the supervisor (Graen et al, 1977) at the same time satisfaction with the supervisor is significantly correlated with self-reported satisfaction (Riggio & Cole, 1992).

Even when researchers and practitioners approach this topic from different perspectives and for different reasons, the consistent result is that it's beneficial to everyone involved when people within an organization are happy (Weizbord, 1991). In answer to the question of whether quality of work life really works better, Cohen responds with a resounding "yes" (1979).

Studies identify another interesting factor influencing subordinate job satisfaction. Supervisor influence in the organization directly impacts subordinate satisfaction (Pelz, 1952). Subordinates are more satisfied with supervisors who possess upward influence (Jablin, 1980). While these relationship aren’t examined in my study, they are another general argument in favor of TQM implementation.

*What Do Subordinates Expect of Their Supervisors?* Research indicates some universal expectations of supervisors. In a study identifying these behaviors, Wemimont (1970) identified seven major areas: freedom to do their job their own way; communication about the job and performance feedback, honesty and fairness; genuine interest in the subordinate and the subordinate’s work; leadership; confidence and support; and capability.
I modified a questionnaire used by the Merit Systems Protection Board (MSPB) to prepare a report for the President and Congress on the status of federal first line supervisors. This questionnaire is based on work conducted by Office of Personnel Management (OPM) (U.S. Merit Systems Protection Board, 1992). The Management Excellence Inventory, the result of four years of development by OPM, divides managerial responsibilities into two categories: management functions and effectiveness characteristics. Management functions, the "what," are the content of the responsibilities. Effectiveness characteristics, the "how," are the skills, attitudes and perspectives. This work of OPM is well documented and considered adaptable for use in other organizational settings (Flanders & Utterback, 1985).

There have been several attempts to identify thorny issues within the supervisor-subordinate relationship. A survey of primarily public sector subordinates identified characteristics poor supervisors possess: poor organizational performance, poor decision-making skills, poor communication skills, poor professional interpersonal relationships, undesirable personal characteristics (Brown, 1964). A survey of accountants revealed no surprises: failure to provide support, failure to provide an environment conducive to the exchange of ideas; and failure to be honest, truthful and straightforward were the main issues identified (Hammer & Gavin, 1983).

Perception

Responses gathered in this study are ratings of supervisor task performance, these ratings are based on the perceptions of the subjects. Supervisor and subordinate evaluations of the supervisor have to be taken at face-value: no criteria for measurement were provided. This freedom raises a number of interesting issues.

When an organization is trying to change, as the experimental group hoped, perceptions can be pivotal to success or failure. Covin and Kilmann (1990) identified factors influencing change by examining the perceptions of people within organizations attempting to implement change. Positive influences include: visible support and commitment; encouraging employee participation; a high degree of communication; and a reward system that supports the necessary changes. Negative influences include: lack of management support; inconsistent actions by key managers; unrealistic expectations; lack of meaningful participation; poor communication; and misplacement of responsibility.
How Does Perception Work? Descriptions of the behavior of one person by another tend to be subject to biases, including a tendency to attribute favorable behaviors or motives to liked persons and unfavorable behaviors or motives to disliked persons" (Vroom, 1982, p. 113). The perception of other people is somewhat arbitrary. They may be perceived as opportunities, threats or irrelevant without any knowledgeable basis. A person’s perceptions "are limited by the perspective or position from which he does the observing" (McCall & Simmons, 1966, p. 104).

Are Self Evaluations Accurate? Supervisory self perceptions aren’t significantly related to observed supervisory behavior (Gross, 1956). "People will engage in a wide array of mental gymnastics to dismiss discrepant feedback" (Swann, 1983, p.44). "Inflated self-appraisals have been found in groups of all types, both blue collar and white collar. In general, self-appraisals tend to be a little more modest at lower levels than at high levels" (Meyer, 1980, p.293). Anonymity appears to be an important factor in self evaluation accuracy (Mabe & West, 1982).

Are Evaluations by Subordinates Accurate?
Bernardin (1986) provided the most comprehensive discussion I could find on this issue: identifying potential drawbacks and benefits to the use of subordinate appraisals, concluding they are a valuable source of information. Evaluations from subordinates tend to be more similar to evaluations from superiors than to self evaluations (Mount, 1984). This is another argument supporting the use of evaluations from subordinates. Since a supervisor also has the role of subordinate in another relationship, there is no reasonable logic to explain why their views are credible in one direction and not the other.

Until this study, I never heard of subordinates evaluating their supervisors, although I often wished for an opportunity to have my opinions heard. Subordinates should be provided a vehicle for confidentially evaluating their supervisor. I think it’s in an organization’s best interest to learn what is going on. Since supervisory responsibilities are a large portion of a supervisor's duties, a critical ingredient in their contribution to the organization, subordinate input should be included in the official performance evaluation of the supervisors. The measures should be defined, including the scale. Suitable topics include: fairness and consistency, task performance, organizational effectiveness and overall impressions. I haven’t seen instructions for the preparation of an evaluation format for soliciting the opinions of subordinates. I had these ideas before returning to school. They are reinforced by what I’ve learned in classes and from my study.
Having the subordinates evaluate their supervisor's task performance is a twist on the norm. The annual performance review, supervisor judging subordinate, is an organizational ritual. Soliciting subordinate input isn't a new idea, but it's controversial. Many human resource managers recommend gathering upward appraisal information and presenting it to managers in ways that won't reveal "what any individual said" (Nicholas, 1992, p. 17). Some managers vehemently oppose the notion entirely. Public sector employees are reluctant to endorse the collection of subordinate views and are emphatic in their position that these responses not be used to officially evaluate them (McEvoy, 1990).

There is a lot of literature addressing self evaluation, often contrasting self evaluation with evaluations from other sources. Matching supervisors and subordinates is a rather uncommon methodology, although similar schemes were used successfully by Tharenou and Lyndon (1990) and Eisenberg et al (1984). In my study, the "self" is the supervisor. Typically, the "others" are the subordinate or subordinates of the supervisor, the supervisor of the supervisor, also called the superior; and peers of the supervisor. In my study the "other" is a paired subordinate.

Issues of fear arise often. A subordinate evaluating a supervisor may feel fear that they will endure retaliation if they report some unwholesome truth. A supervisor may feel fear about being honest on a self evaluation if they report a weakness.

*How Do Evaluations by Self and Subordinate Compare?* Overwhelmingly, previous studies show that supervisors tend to rate themselves higher than subordinates rate their supervisors. Generally, self evaluations tend to be most lenient. The exception to this finding is from Fox and Dinur (1988). In this study, exceptional junior officers were selected to participate in special Army training. Each subject was asked to evaluate his performance. These ratings were compared with ratings from superiors. The two sets of ratings were remarkably consistent: the subjects performed at the highest levels. For me, this study has no value. After selecting candidates for the school, supervisors would be reluctant to admit their subordinate performed less well than the other students. It's not surprising to me that people who are constantly told they are special will believe they are special. So there is really no room for comparison when supervisors and subordinates both intend to give the highest evaluations.
A variation of the comparison between self and other evaluations examined whether supervisors accurately perceive how their traits are perceived by their subordinates. Infante & Gordon (1979) found that supervisors underestimate how favorably they are perceived by subordinates.

Michael Mount (1984) conducted the most definitive studies comparing evaluations by self with evaluations by others. He found self, subordinate and superior ratings of supervisory performance are significantly related to subordinate satisfaction with supervision.

Leniency is the extent to which one evaluation is higher than evaluations from other sources (Thornton, 1980). Leniency is present when the levels of performance ratings obtained from different sources are significantly different (Mount, 1984). Self evaluation tends to show more leniency than superior, subordinate or peer evaluations (Thornton, 1980) which have good correlation between them (Wohlers & London, 1989). Subordinates tend to describe their supervisors more autocratically than supervisors report themselves to be (Jago & Vroom, 1975). If people know in advance they will be receiving feedback confirming their self conceptions, they are especially attentive (Swann, 1983).

Individuals rate themselves higher than they are rated by comparison groups (Thornton, 1980). Even when there is significant agreement among multiple subordinates of a single supervisor, there is little agreement with the self evaluation of the supervisor (Jago & Vroom, 1975). Self awareness is the degree to which individuals see themselves as others see them (Wohlers & London, 1989, p. 236). "Individuals have a significantly different view of their own job performance than that held by other people" (Thornton, 1980, p. 268). Subordinate descriptions are significantly lower than self descriptions (Jago & Vroom, 1975) but there is correlation between them (London & Wohlers, 1991).

Supervisors don't accurately perceive how their traits are perceived by their subordinates. Infante & Gordon (1979) asked supervisors to predict the evaluations they would receive from their subordinates. Supervisors felt they have more control than subordinates perceive and supervisors tended to underestimate how favorably they are perceived by their subordinates.

Agreement between supervisor and subordinate regarding the meaning of certain mutually experienced events and situations depends on the quality and openness of their relationship (Graen & Schiemann, 1978). For example, supervisors and subordinates often have different views about
their communication. "Subordinates feel they must take seriously the communication interactions they have with the boss, but the boss does not have to reciprocate the same degree of attentiveness and favorable reaction" (Porter & Roberts, 1976, p. 1575). This difference in perceived weight assigned to a conversation has implications beyond the internal relationship of the pair. The degree of mutual understanding is positively related to performance (Bernard et al, 1961). The Eisenberg et al (1984) study determined supervisors and subordinates at all levels consistently misperceive their communication with each other.

Is There Value in Having Subordinates Rate Their Supervisors? Yes, but I don't believe useful evaluations can occur in an unenlightened organization. I don't think it's possible for either the supervisor or subordinate to recognize and articulate the other's activities unless they have some rudimentary understanding of what is desirable from the organization's perspective, implying some degree of personal and professional development.

Subordinate satisfaction with supervisory behavior isn't significantly related to observed supervisory behavior (Gross, 1956). According to Cooley (1986), performance feedback and goodness of job instructions are two of the best predictors of perceived supervisory performance. Not surprisingly, he also found that performance feedback and organizational procedures and practices predict satisfaction with a supervisor. The questions on my instrument include these less visible tasks.

Merit Systems Protection Board (MSPB) prepared their 1992 report to the president and Congress on government supervisors using data collected from subordinates, superiors and supervisors. One of the concerns noted in this report is the discrepancy between the subordinates' views and the supervisors' views. The conclusions regarding the appropriateness of these data sources states "a well-rounded assessment of a supervisor's quality or effectiveness should include input from all three levels since each provides a potentially unique perspective" (1992, p.2).

The debate about the validity of subordinates' appraisals isn't new. Since subordinates are in a unique position, they can capture unique observations (McEvoy, 1990). "Properly implemented and monitored subordinate appraisals can be a meaningful source of feedback for all involved supervisors, subordinates, and the organization at large" (Bernardin & Beatty, 1987, p. 63). Subordinate's feelings of vulnerability must be protected (Osborne, 1990).
Arguments used by opponents to the use of supervisor appraisals from subordinates include: subordinates probably lack ability and information needed to provide valid appraisals; subordinates may be too strict in their ratings of supervisors who push hardest; and subordinates may slant or inflate their appraisal in an attempt to gain influence. There is some evidence indicating supervisors could focus too heavily on pleasing their subordinates, undermining supervisor authority or turning the evaluation into a popularity contest (Bernardin, 1986). Some proponents argue that overcoming subordinates' fears is a greater obstacle in the gathering of true appraisals (Bernardin, 1986; Osborne, 1990; Nicholas, 1992; London & Wohlers, 1991). Weinrauch (1975) obtained a surprising result when he studied the honesty of subordinate appraisals: subordinates expecting a financial reward soon rated their supervisors higher.

Recognizing that subordinates can provide valuable insight, supervisors are asking for feedback from their subordinates with increasing frequency (Nicholas, 1992). Even public sector supervisors allow that subordinate appraisals can be useful for guiding the development of the supervisor (McEvoy, 1990).

Despite broad enthusiasm for subordinate appraisal of the supervisors, many supervisors are reluctant to allow subordinate perspectives to become part of the performance reviews they receive. Some human resource managers fear that supervisors will view subordinate appraisal as sufficient threat that will adversely affect recruitment and retention (Bernardin, 1986). The Management Excellence Inventory, the precursor to my instrument, was developed to help assess and identify areas of development for federal government employees, not to provide performance appraisals (Flanders & Utterback, 1985).

In my opinion, the subordinate appraisals of the supervisor should be reviewed by the supervisor’s superiors. How the supervisors conduct their supervisory responsibilities is a critical ingredient in how they are contributing to the organization. The responses can be used to determine supervisor development efforts, but should also be used to determine the rewards a supervisor receives.

Training
Research comparing the introduction and implementation of various quality theories into real working environments concluded there is no significant difference within an organization when one approach is chosen over another. Irrespective of whether Deming, Juran or any other "gurus" are selected as the medium for teaching TQM to organization members, the results are the same. An
interesting observation is that Tom Peters initially instills the greatest degree of excitement and enthusiasm (Stieber, 1991).

Much of the literature differentiates between training, employee development, and education. After carefully reading the different definitions, I cannot see consistency in the application of the assorted terminology nor can I describe the differences. For my study, I use only training. All my discussions of training evaluation are limited to conceptual training. They don't apply to concrete skill acquisition, procedural demonstrations or production situations.

Training is a common theme when quality and people are discussed. Follett identifies three ways to deal with conflict: domination, compromise and integration. Without reservation, she considers integration to be the only reasonable route. Integration occurs when two or more desires have a place in the solution and neither side has to sacrifice anything. She reasons that the biggest obstacle to integration is our lack of training in cooperative theory (Follett, 1940). McGregor considers self-development a basic need and maintains "there is no sound basis for expecting an individual to sacrifice his personal goals or growth for the organization (1960, p. 24). Drucker considers knowledge a prerequisite for success. He calls knowledge workers the "new majority" and asserts "it makes absolutely no difference to their economic or social position whether they work for a business, a hospital or a university" (Drucker, 1989, p. 25).

Dr. Deming devotes a great deal of thought to training. Of his fourteen points, two are devoted to training. Point six addresses foundations of training. Point thirteen elaborates on the need for continual education and self-education. Employees are competitive assets and must have the skills and knowledge to contribute to the organization. With vision and training, his number one deadly disease, lack of constancy of purpose, can be vanquished (Deming, 1986). The objectives of the experimental group's training initiative are coincident with Dr. Deming's goals.

Learning. There is no consensus on learning. I don't agree with Kirkpatrick that learning is "the principles, facts and skills which were understood and absorbed by the conferees. In other words, it does not include on-the-job use of these principles, facts and skills" (1976, p. 18-11). I think learning involves the acquisition of skills for diagnosing situations and applying appropriately the knowledge learned. Learning is an increased knowledge of facts, concepts and techniques. Changes in expressed attitudes, including acceptance of the training content, demonstrate learning (Watson, 1979).
Training Evaluation. Kirkpatrick dominates practitioner's literature while Solomon's Four Group Design (1949), explained in Methodology, Chapter 3, is preferred by researchers. Even Kirkpatrick alludes to Solomon's design when he recommends using a pretest and having a control group when giving a post-test to measure learning.

Kirkpatrick (1976) is recognized as the practitioner's authority on training evaluation, referenced in many books and articles and apparent in those not naming him. His original work was directed toward in-house training efforts, but may be applied to any training situation. He breaks evaluation into four steps: reaction, learning, behavior and results. I can see the logic behind the four steps, but I am uncomfortable with the sequencing implied in Kirkpatrick's terminology. Many details are ambiguous, leaving too much room for interpretation. Alliger and Janak (1989) wrote of his model, "its easily adopted vocabulary and a number of (often implicit) assumptions, can tend to misunderstandings and overgeneralizations" (p. 331). His impact on training evaluation in the workplace cannot be questioned. The procedures and outlines presented in numerous articles directed at practitioners reflect tenets of his theory.

I like Kirkpatrick's recognition of more than one level of evaluation. But I prefer Wolfe (1973) in some respects. Evaluation can and should test four areas: a reaction from the trainee; the amount of learning and attitude change; the difference in the way the trainee behaves as a result of the training; and any improvement in the work unit's performance. Wolfe recognizes the difficulties inherent in the evaluation of training and suggests experimental program evaluation, a methodology reflecting the Solomon four-group design with a long-term perspective. I think collecting data must be carefully planned. Quizzing people is rather demeaning, implying a punitive attitude that instills fear. Asking feelings about the training experience doesn't give a good sense about the effectiveness or efficiency of the training. Excessive data collection unnecessarily burdens the people gathering the data. Gordon (1991) defers to Kirkpatrick in theory, but demonstrates an understanding of the real-world's time and budget constraints.

Neither Kirkpatrick nor Wolfe answer all my questions. I appreciate the different types of evaluation Kirkpatrick elaborates on, but I am left wanting more. I would rather learn the factors that might influence the evaluation step than read examples of each step. I think his theories can be addressed more succinctly if presented in the manner of Watson or Gordon. Wolfe is logical, but I
have trouble distinguishing amount of learning and attitude change from differences in the way the trainee behaves as a result of the training.

I don't agree with Kirkpatrick that all training MUST be evaluated. The application of the training evaluation must first be considered. This step is particularly important when dollars are being assigned. There is a cost to collecting data and there should be an intended use. Probably, there is a real use for evaluation of most training, but the use should be identified. While Watson (1979) doesn't contradict Kirkpatrick, his view of evaluation is a bit more tempered. Some level of training can be considered part of the employee compensation system. Only needed data should be collected; all data have a cost. There are three approaches to measuring learning: nonsystematic collection of data, systematic collection of indicators and experimental. Subjects may balk at being the focus of scrutiny and may be insulted if tested. The experimental approach involves pre- and post-data collection. Whereas evaluation of training can be straightforward in production operations, where results and output can be measured, it's difficult to do when concepts, problem-solving and ability to work with others are the point of the training. Watson cautions, "Favorable reactions to a program and to those who presented it are indications that participants liked the experience, but positive reactions are negligible assurance that any learning actually occurred" (1979, p. 288).

I don't agree with Kirkpatrick's belief that a person can learn something and not apply that knowledge to their life. His distinction between "learning" and "behavior" confuses me. While I can see that a practitioner might have to use different measures for learning or behavior, I consider the behavior to be an extension of the learning; unless a trainee has no opportunity to demonstrate what has been learned (as with a skill), applications of the learning should be apparent in subsequent behavior. Regurgitating theory and terminology doesn't mean someone has learned what they were taught. For example, a child may be taught repeatedly that biting isn't appropriate behavior. When questioned by an adult, that child can list the reasons and alternatives. Yet, biting behavior continues. I don't think the child has learned polite behavior, while Kirkpatrick would disagree, saying the child has learned but doesn't implement what he knows.

To measure learning, Kirkpatrick recommends evaluation using statistical analysis of pre- and post-data collection involving the person, superiors, subordinates and peers. The post-test should be at least three months after the training. Again, I think the use for the information must be considered when the evaluation is considered before the evaluation process is defined. According
to Mindak & Anderson (1971), behavior changes are rare. The difficulty in observing a change is compounded by the need to select the right measures. A delayed test design, with pre- and post-testing is recommended. Results can be displayed as changes. They caution that attitude changes are short-lived. For my study, subordinates and supervisors participated in training and both were asked to provide data; the post-test was scheduled for approximately two months after the formal training program was completed, allowing time for the subjects to practice what they were taught.

I wish it were possible to convert training to dollars, but where I've seen examples, there are assumptions and approximations which remove any confidence I have in the numbers. Others disagree. Carnevale & Schulz (1990) devote a great deal of energy to converting all training efforts to dollars. They reason that training, like any other investment, must justify itself with a reasonable return on investment and focus on Kirkpatrick's fourth step. By going to great lengths collecting data, the evaluation process can be costly. Making sweeping assumptions to perform calculations reduces the credibility of the results. Where figures are available, it's reassuring to see the return on a training investment, but this isn't always convenient. Watson (1979) remembers that the people participating in the training are individuals with feelings. He recognizes that training opportunities may be part of an organization's reward system. He understands the potential for complications in the collection of evaluation data, and cautions that the requirements should be identified beforehand.

My Conceptual Model
The conceptual model combines all the topics outlined in this Chapter. The introductory TQM training selected by the experimental sub-organization was a program endorsed by Dr. Deming. I've tried to show how the concepts inherent in TQM have a long history. The areas I've addressed in my literature review relate to why I hypothesize introductory TQM training will decrease the differences between supervisor and subordinate perceptions of supervisor task performance. Since the supervisor-subordinate relationship is important to the organization, I focus my attention in this study on this relationship. I chose to measure perceptions using an instrument directed at supervisor task performance. Supervisors generally evaluate themselves more leniently than their subordinates evaluate them.

After all these foundations are laid, I'm looking at the evaluation of a training program an organization invested in to improve while facing tighter financial constraints. There are different theories on training and training evaluation. I believe the actions of the supervisors should reflect
what they've learned, so I consider perceptions of supervisor task performance appropriate. I wasn't looking at organizational improvements because of the training. I'm interested in the supervisor's and paired subordinate's perceptions of the supervisor's task performance. The longitudinal change in supervisor task performance is virtually ignored. I focus on the difference between the two perceptions as measured by the assigned ratings.
Chapter 3
METHODOLOGY

In Chapter 3, I describe the research methods I employed to examine how introductory TQM training affects supervisor and subordinate perceptions of supervisor task performance within established supervisor-subordinate relationships in an administrative sub-organization of an American land-grant university. I'm interested in measured differences within the pairs.

The following research sub-questions are addressed:
1.) How do the experimental and control groups differ after the treatment?
2.) How do the experimental and control groups differ prior to the treatment?
3.) Does the measurement instrument have an effect on paired response differences?
4.) Are differences between the supervisor and paired subordinate affected by the treatment?
5.) Can the effects of the treatment be observed within the experimental group?

Research Design
Measurements of supervisor and subordinate perceptions of supervisor task performance are taken pre- and post-training. During the pre-test and the post-test, both supervisors and subordinates rate the supervisor's task performance. These ratings are based on the perspective of the rater. The differences in the supervisor's perspective of the supervisor's task performance and the subordinate's perspective of that supervisor's task performance for established supervisor and subordinate pairs are compared.

For both the experimental and control group, the measurements are gathered concurrently. All experimental group subjects participate in introductory TQM training. Participation in the training affects the perceptions of the supervisor and subordinate and may affect supervisor task performance. My research design model is presented in Figure 3.
Experimental Design

A Solomon Four Group Design. A Solomon four group design is a classic experimental design for testing the effects of a treatment while controlling for the effects of the measurement instrument (Campbell & Stanley, 1963). With this design, the impact of the pre-test instrument can be measured, the end results can be compared, and pre-test subject differences can be observed. The
traditional presentation of this design is in Table 1 below (Solomon, 1949). This arrangement makes it easy to see what each group does. My experimental group subjects participating in both the pre-test and post-test are the Experimental Group. My control group subjects participating in the pre-test are Control Group I. My experimental group subjects participating in the post-test only are Control Group II. And my control group subjects participating in the post-test only are Control Group III.

<table>
<thead>
<tr>
<th></th>
<th>Experimental Group</th>
<th>Control Group I</th>
<th>Control Group II</th>
<th>Control Group III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Training</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Post-test</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

*My Design.* I've rearranged Solomon's table slightly. My experimental group includes both groups receiving training (Experimental Group and Control Group II) and my control group includes both groups not receiving training (Control Group I and Control Group II). I've also rotated the table so the groups are on the vertical axis and the tests are on the horizontal axis. I think this arrangement shows the passage of time, particularly useful when there are more than one post-test. My design is found in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>PRE-TEST DIFFERENCES</th>
<th>POST-TEST DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL</td>
<td>pair₁, pair₂, pair₃</td>
<td>pair₁, pair₂, pair₃</td>
</tr>
<tr>
<td>GROUP</td>
<td>pair₄, pair₅</td>
<td>pair₄, pair₅</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td>pair₆, pair₇, pair₈</td>
<td>pair₆, pair₇, pair₈</td>
</tr>
<tr>
<td></td>
<td>pair₉, pair₁₀</td>
<td>pair₉, pair₁₀</td>
</tr>
</tbody>
</table>
For both phases of the study, the pre-test and the post-test, the tests for the two groups were administered simultaneously. The subjects in each group were randomly divided so half the pairs in each group participated in the pre-test. All the pairs participated in the post-test 3.

Participants responded by marking responses on opscan forms. The measurement instruments for the two groups differed slightly, with a few items altered to be consistent with terminology for each office. Supervisor and subordinate versions differed slightly so all items related to the supervisor from the appropriate perspective. Details of these variations are provided in my discussion of the measurement instrument.

Previous research (Flanders & Utterback, 1985; U.S. Merit Systems Protection Board, 1992) identified the tasks Federal Government supervisors perform. These 118 tasks are reflected in my instrument. Individual questions are grouped into subscales reflecting different dimensions of twenty-one concepts 4 of supervisor task performance.

Population
The population of interest for this study is American land-grant university administrators. The sample consists of two naturally occurring sub-organizations at the same university. Each of these sub-organizations is relatively autonomous within university guidelines. Subjects are paired supervisors and subordinates with well-established relationships; i.e., employees new to the sub-organizations are excluded from the study. All subjects are permanent, full-time employees of the sub-organizations.

The two sub-organizations are similar in size but aren't identical in structural arrangement. I ignore the internal hierarchies. I'm not interested in the relationships between the supervisors.

Both sub-organizations are arranged so people work closely with a limited number of people. Primarily these supervisor-subordinate relationships are paired, but there are a few supervisors who have more than one subordinate. The responses from each supervisor are considered only once so where a supervisor has more than one subordinate, only the responses from a single subordinate are considered in my analysis. I'm not interested in the relationships among subordinates. My analysis focuses on supervisor-subordinate pairs. Each supervisor has only a single subordinate relationship in my study; the responses of the second subordinate are irrelevant.
except for instrument reliability calculations. If more than one subordinate responded fully to the survey, the response of only one subordinate was used; that subordinate was selected randomly.

Measurement Instrument

A generalized copy of the measurement instrument is in Appendix A 5.

My Measurement Instrument. The first part of my instrument is 118 questions from a U.S. Merit Systems Protection Board (MSPB) questionnaire. Each question asks about supervisor performance on one of the tasks identified by the U.S. Office of Personnel Management (OPM) as part of a federal supervisor's job (Flanders & Utterback, 1985). The 118 questions are clustered into twenty-one subscales. Each question in a subscale is a different facet of the subscale concept; there is no repetition. The rating scale for these items is at the top of each page. My options for performance level are: doesn’t do this task, cannot do this task at an acceptable level, can do this task at a barely acceptable level, can do this task at an acceptable level, can do this task at an above average level, can do this task exceptionally well, and can’t judge.

The last part of my instrument is composed of a few general-impression questions. Depending on whether the instrument was intended for the pre-test or the post-test and the experimental or control group, the number of general-impression questions varies from six to ten. The rating scale for most of these items allows six options.

My Modifications. The MSPB questionnaire has three parts to every task-related question, mine has one: how do you (or your supervisor) perform a given task. The MSPB has six options for performance evaluation, I have seven. (I added doesn’t do this task.) I also changed a few key words so the measurement instrument would be consistent with office lingo (e.g., supervisor or faculty member and subordinate or support staff). For example, the instrument for supervisors asks about your performance while the instrument for subordinates asks about your supervisor’s performance. The instrument for the experimental group uses the term supervisor while the instrument for the control group uses the term faculty member.

That my measurement instrument contains no demographic questions was a deliberate decision. Identifying information might have harmed the study. Subjects could have been reluctant to provide honest unfavorable responses if there was a possibility they would be identified. Subordinates were being asked to rate their supervisor’s performance. Supervisors knew some of
their peers would be looking at the responses. I assured confidentiality with a coding scheme and by omitting descriptive questions from the instrument.

There are two versions of the measurement instrument for each test for each group: one for supervisors and one for subordinates. The 118 tasks are addressed in the same order on both versions. Both versions ask for details of supervisor performance. The wording differs slightly for the experimental and control groups (e.g., supervisor or faculty member and subordinate or support staff) to be consistent with the terminology used within each sub-organization.

The Questions. Part of my study is inseparable from the training evaluation conducted by the experimental group training facilitator. For research purposes, I used the same measurement instrument for both the pre-test and post-test. The first 126 questions of my instrument are identical for the pre-test and post-test. A few questions of interest to the experimental group training facilitator were added to the post-test, for a total of 129 questions. By putting these questions at the end, the questions of interest to me shouldn't be influenced.

The Audience. My measurement instrument is directed at identified supervisor-subordinate pairs. A coding scheme allows me to identify who is a supervisor and who is a subordinate, whether a subject is from the experimental or control group, and pre-test and post-test responses.

Normative Data. My measurement instrument, built on the work of MSPB and OPM, has no normative data.

The MSPB questionnaire, the basis of my measurement instrument, had a 62-65% return rate for the survey conducted for the 1992 Report to the President and the Congress of the United States. From this survey, the most important tasks for a first-line supervisor’s job were identified by MSPB. See Appendix B 9.

More than five years of research conducted by OPM, the basis of the MSPB questionnaire, had three stages: a random survey of federal managers, open-ended interviews with subjects from ten agencies, and job analysis. Despite similarities in supervisor tasks, their relative importance varied depending on whether the supervisor was an executive, a manager, or a lower-level supervisor. The OPM researchers were satisfied that their findings were consistent with other managerial research.
Reliability Data. Neither my measurement instrument nor the MSPB questionnaire, the basis of my measurement instrument, have reliability data. OPM measured internal consistency of several administrations of their Management Excellence Inventory tool, the precursor to the MSPB questionnaire, using Cronbach's Alpha. Results ranged from 0.6 to 0.8.

Pilot Study
I didn't conduct a pilot study. An opportunity to carry out research for my thesis was made available to me. I had to act immediately, with only a few weeks to prepare for the pre-test. I applied a questionnaire already given to thousands of subjects. The U.S. Government has such confidence in this questionnaire that it was used to gather data for a report to the President and Congress (U.S. Merit Systems Protection Board, 1992). This report was being informally circulated in the Industrial and Systems Engineering Department at the time the experimental sub-organization was planning to begin TQM training.

Even if I had time for a pilot study, it would have been difficult to find subjects parallel to the subjects used in my study. I used two university sub-organizations in their entirety. These sub-organizations are comparable, not identical, in size and structure. Although TQM activity was occurring on the university campus, the control group hadn't officially been exposed to TQM training. Their top management agreed to delay the introduction of TQM to the sub-organization until after the post-test. There isn't any way to determine the extent of informal exposure subjects in either sub-organization had before the start of my study. The literature suggests certain academic areas are more likely to either incorporate or not incorporate TQM in their operations. The two sub-organizations participating in my study seem to fit this trend.

The original MSPB instrument was based on years of OPM research on federal government employees. I considered state government operations similar to federal government operations. Since all my subjects fill administrative positions, I considered their job responsibilities comparable to any other state government administrator. Based on this logic, I felt comfortable using the instrument with minor modifications to collect data for my research. That the originators of the model behind the instrument were satisfied their findings were consistent with other research added to my confidence in the instrument.
Subject Confidentiality

Collecting the Data. My job experience conducting a large survey for the U.S. Government had already sensitized me to some of the fears people have when being surveyed about their jobs. Because the two groups are small enough that a person could be linked with a response, I've been careful to protect identities. I assured confidentiality both to reassure the subjects and to solicit honest responses.

The data collection procedure ensured no responses would be linked with any individual. Supervisors and subordinates received their surveys together but returned them separately. Envelopes were provided to provide the option of sealing the responses from prying eyes. Collection points weren't monitored lest the subjects worry I might associate a response with an individual.

Coding. To link subordinates with their supervisor and maintain anonymity, identifying numbers were randomly assigned to each pair. The numbering scheme allowed me to differentiate between supervisor and subordinate responses and experimental and control group responses while tracking the pairs over time. To preclude any confusion with numbers, assigned numbers were recorded and subjects sealed their numbers in envelopes which were collected for safe keeping.

The codes were on the opscan forms at the time of distribution for first-time participants. The supervisor-subordinate relationships of interest to me were defined while I was planning, so I knew how many pairs and trios I had. For the control group, this step was necessary since some of the subordinates support more than one supervisor yet have a primary supervisor. Only enough surveys for each established relationship were included in the distribution packets. Second-time participants had to complete their codes on the opscan forms. (The supervisor/subordinate verification codes and the last digit of the pairing codes were already marked.) To avoid future coding confusion, code numbers were sealed in envelopes marked with the subject's name. The envelopes were collected and locked in a safe. Only the subject was allowed to open the envelope. This ensured consistency of the coding process over time.

Within the Thesis. There's no possibility any response can be associated with any individual. All pairs are assigned anonymous subscripts, allowing me to precisely describe what is happening in the analysis without jeopardizing the privacy of any subject. These subscripts are the ones I use in Table 2, as explained in note 2. I'll use the same subscripts in my Results chapter.
Treatment

Over the course of three months, the experimental group participated in a series of almost-weekly training sessions, primarily within the confines of their regular environment. Typically, a training session started with a film from the Deming Library and was followed by open discussion. Topics for discussion were suggested in the pamphlet accompanying each film. The experimental group training facilitator solicited comments from each participant. Attendance at these meetings was mandatory for all experimental sub-organization full-time employees. There was an effort to hold these meetings on a regular basis, but a few weeks were missed because of scheduling conflicts. The training schedule is in Appendix C.

I didn't attend the training sessions. I'm not part of the sub-organization, so I might have inhibited the freedom of discussion. Only sub-organization members attended the training. All discussions were facilitated by the same experimental group training facilitator.

Data Collection

The Survey. My survey has three parts: instructions, measurement instrument, and opscan form. Instrument wording is tailored for supervisor or subordinate and experimental or control group. Opscan forms are coded for both supervisor or subordinate and experimental or control group. An additional code was used to verify supervisors and subordinates completed the correct form.

The Procedure. Similar data collection procedures were used for the two groups. Subjects from the two groups never met together for the purpose of this study. The experimental group training facilitator coordinated the participation of the control group sub-organization. I collected all parts to the returned surveys.

All potential subjects were introduced to the study shortly before the treatment was scheduled to begin and approximately one month after the formal treatment was completed. A meeting was held to explain the procedure and distribute survey packets to the experimental sub-organization. The control sub-organization attended a similar meeting. Subjects were asked to complete the surveys individually and refrain from discussing responses with each other. The importance of proper coding was stressed. Collection points were set up in each office and emptied daily. Returned responses were tracked to ensure all surveys were collected.
Interview with the Experimental Group Training Facilitator

Shortly before beginning my analysis of the responses, I conducted a tape-recorded interview with the experimental group training facilitator. Our conversation is summarized in Appendix D10.

Analysis Methods

Focus of Analysis. Even using subscales, no interpretations are obvious. A manageable examination of the responses requires some logical focusing without bias. To determine where to concentrate my attention, I conducted a secondary survey after the primary survey. The results of this survey determined the subscales I use for my primary survey analysis.

The secondary survey instrument is a list of fourteen of the twenty-one subscales from my primary measurement instrument. I selected these subscales based on impressions I gleaned from interviewing the experimental group training facilitator. I believe each of these subscales is of some importance to the supervisor-subordinate relationship. Respondents were asked to rank in descending order the ten they consider most important to the relationship. A copy of the secondary survey instrument is in Appendix E.

The volunteer sample participating in this secondary survey is people formally associated with the university where the experiment was conducted and having permanent, full-time work experience. None of the respondents are subjects in the primary survey. Summer work, part-time work, teaching assistant positions, research positions and volunteer positions don't meet my definition of permanent work experience. None of the respondents are undergraduate students.

On a day when the university was open but not in session, I wandered around campus collecting random responses. I collected responses from people associated with different departments and academic areas. I hoped the quiet environment would make people conducive to participation. A total of twenty-one people provided input.

Response Analysis. My statistical analysis consists of a series of independent t-tests of the differences between supervisor and subordinate responses.

For each pair, the difference between supervisor and subordinate responses for each of the 118 task items was to be calculated and stored in arrays structured like Table 2. I expected to be able to
examine each question, paying particular attention to the fourteen most important tasks for first-line supervisor's jobs. See Appendix B.

The analysis is done primarily on a subscale level, necessitated by the large amount of data that couldn't be compared. Each subscale is assigned a calculated score for each subject. The differences in subscale scores are compared in the t-tests.

**Threats to Validity.** Because I used natural groups, I face some special problems with the interpretation and generalization of the results. The opportunity to collect data in a real environment rather than a controlled laboratory setting offsets some of these reservations: the data are real, not simulated.

Despite the fact that two sub-organizations are measured, the sample size is low. Small sample size reduces the power to detect differences. The results possibly could be reversed from the true population because of sampling fluctuations (Campbell, 1988).

Campbell and Stanley (1963) identified eight potential sources of internal validity and four potential factors jeopardizing generalizability. When a true experiment uses the Solomon Four Group Design, all the sources of internal and external validity are controlled except two: the interactive effects of selection biases and the experimental variable and the reactive effects of experimental arrangements. I don't have a true experiment since I didn't have random assignment to treatment. Undefined variables couldn't be controlled; the groups aren't statistically equivalent. Without random assignment, there is no generalizability of the results to the population.
NOTES

1. My research can be classified as a quasi-experimental field study. I'm trying to determine whether introductory TQM training makes the perceptions of the supervisor and the subordinate, reflected in the ratings assigned, more similar than they would be without introductory TQM training. I'm not attempting to solve an existing problem.

2. Whenever I refer to specific subjects, I use numeric subscripts and letter subscripts to denote experimental and control group pairs respectively. Pairs one through five (pair_1 through pair_5) and a through d (pair_a through pair_d) are the subjects completing both the pre-test and the post-test. Pairs six through ten (pair_6 through pair_10) and e and f (pair_e and pair_f) are the subjects completing the post-test only. This scheme is followed throughout my thesis.

3. Supervisor-subordinate pairs in both the experimental and control groups were randomly assigned pre-tests. At the time these assignments were made, both groups had approximately half their people participate in the pre-test. Attrition during the five-month interval between the pre-test and post-test resulted in an unbalanced division of the control group.

4. The subscales are the effectiveness characteristics and management functions described by Flanders and Utterback (1985). Effectiveness characteristics (the how) include: broad perspective, strategic view, environmental sensitivity, leadership, flexibility, action orientation, results focus, communication, interpersonal sensitivity, and technical competence. Management functions (the what) are: external awareness, interpretation, representation, coordination, work unit planning, work unit guidance, budgeting, material resources administration, personnel management, supervision, work unit monitoring, and program evaluation.

5. There are eight versions of my measurement instrument: pre-test experimental group supervisors, pre-test experimental group subordinates, pre-test control group supervisors, pre-test control group subordinates, post-test experimental group supervisors, post-test experimental group subordinates, post-test control group supervisors, and post-test control group subordinates. The instructions for each version were explicitly tailored to their intended audience. A generalized version of my instrument is in Appendix A.

6. The MSPB questionnaire has three queries for each task: Is the task performed? How important is the task? What is the performance level? Six alternative responses are offered:
cannot do this task at an acceptable level; can do this task at a barely acceptable level; can do this task at an acceptable level, can do this task at an above average level; can do this task exceptionally well; and can't judge.

7. The MSPB questionnaires are completed by second-level supervisors, supervisors and non-supervisors. The responses are compared based on categories, job classifications, and other demographic criteria solicited with the questionnaire; there is no matching of respondents into established work relationships.

8. These are items added to my original measurement instrument by the experimental group training facilitator: Frequency of providing feedback to staff. When you meet with your support staff, how effective are these meetings? Considering all relevant factors, to what extent do you contribute to the accomplishment of the organization's goals and missions?

9. The most important tasks for first-line supervisor's job "were rated by at least 85% of first line supervisors and second-level supervisors as being important to the first-line supervisor's job either to a considerable or great extent." (U.S. Merit Systems Protection Board, 1992, p.10)

10. I chose not to include a script of my interview with the experimental group training facilitator for a number of reasons: the questions were provided to the facilitator prior to the interview so she could collect her thoughts, individuals were mentioned by name in her responses, and our fragmented dialogue would make a script unnecessarily long and confusing.

11. When I began my analysis, I found that some of the responses cannot be compared. For example, where a supervisor says performance can't be judged, how can the difference from the subordinate response be computed for that task? Where a supervisor says a given task isn't performed, how should an exceptional rating from a subordinate be considered? For these reasons, I consider all blanks, 1's (doesn't do this task), and 7's (I can't judge) unusable data.

12. To calculate the subscale score, unusable responses are ignored; if the occurrences of unusable responses are half or more of the questions comprising the subscale, no score can be assigned to that subscale for that subject and no difference can be computed for the pair. If at least half the questions received usable responses, the subscale score is calculated by finding the mean of the usable responses. This mean is the subscale score.
Chapter 4
RESULTS

Chapter 4 consists of the focus of my analysis, primary survey response descriptive statistics, measures of measurement instrument internal consistency, and research hypothesis testing.

Focus of Analysis
Initial analysis of the primary survey results, which was an examination of the results by item, contained some surprises. Based on my literature review, I expected the supervisors to rate themselves more favorably than their subordinates rated them; but this wasn't always what happened. Previous research (U.S. Merit Systems Protection Board, 1992; Flanders and Utterback, 1985) identified the tasks addressed on the instrument as those performed by federal supervisors. My subjects were all administrators at a state university. I assumed the tasks performed for their jobs closely matched the tasks performed by administrators in the federal government. Yet the university administrators felt they either don't perform some of the tasks or they couldn't form an opinion about their performance of these tasks. When paired supervisor and subordinate responses were examined, I could find little correlation or consistency between the two evaluations. Closer examination of these findings are included in Chapter 5.

My Interview with the Experimental Group Training Facilitator. My interview with the experimental group training facilitator was conducted to document the decisions behind the introductory TQM training in that sub-organization. We met after both phases of data collection were completed. Some of the highlights of our conversation are provided.

The experimental group training facilitator isn't responsible for the training and development of the people in the sub-organization, but has a reputation as a champion of quality improvement. For this reason, although other people also attended a Deming Seminar, she was selected to organize and facilitate the training effort for the experimental sub-organization.

The university was facing budget cuts and the sub-organization wanted to get through them in a smart way, without crushing morale. They were thinking of how to restructure the office and responsibilities while improving work at the same time. They had some clear goals for the introductory TQM training: to help people understand the need to look at work from a systems point of view and to help people appreciate how their work affects other people's work. Basically,
the hope was that the training would encourage the office to be more collaborative in working together—to be more customer oriented.

The Deming Library training films were selected for a number of reasons. Some of the sub-organization managers had attended a Deming Seminar and were impressed with his concepts. Sending everybody in the sub-organization would have been too expensive, so an alternative approach was considered. The Deming Library allowed the topics of interest to be specifically selected. The entire library consists of twenty-five tapes. It would have taken too much time for every person to watch and discuss all these tapes. Based on the audience and the needs of the sub-organization, the tapes used in the treatment were purchased. The experimental group training facilitator and top sub-organization managers are satisfied all the topics needed to introduce TQM were presented.

The training format the sub-organization used is the format suggested by the Deming Library. At most training sessions, a video tape was watched by the experimental sub-organization and followed by an open discussion. Each tape comes with a booklet of what to discuss after everyone sees the video. Meetings were mandatory; everyone had to go. Everybody participated, but some were more vocal than others. Everybody had a chance to speak during the discussions.

Some of my questions and summaries of the responses follow. A complete summary of our conversation is presented in Appendix D.

Why did you choose to use Deming's philosophy?
They liked what they saw at the workshop. Based on his definitions, terms, and concepts, he seemed to be more philosophical and more academic than some of the other total quality gurus.

How did you decide to purchase tapes from the Deming library?
She heard about the company that sells the tapes. It was too expensive to send everybody to a Deming workshop. She and the top manager thought the videos were the next best thing.

How did you decide which tapes to use?
She chose the tapes based on what she knows about the staff and what she knows about Deming.

Were your selections limited by a budget?
Somewhat, but a considerable sum was spent purchasing the tapes.
So what was the limit based on?

Time.

How did you decide you had enough training material? Was time the total constraint? Did you have a feeling you covered everything you wanted to cover?

Sub-organization management was playing it by ear, starting with 10 or 12 tapes. The tapes selected turned out to be sufficient background for the staff.

What were the goals of the training?

To help people understand the need to look at work from a systems point of view. To help people appreciate how their work affected other people's work. To encourage the office to be more customer oriented. To help get through some budget cuts in a smart way (thinking through how to restructure office/responsibilities and improve work at the same time).

Were you hoping to see a change in activity, communication or what?

The sub-organization was hoping to see more collaboration in working together to provide better service for customers during the time budgets were cut and expenses were going to have to be cut. Management hoped to gain better service without crushing morale during the budget cut.

The Secondary Survey. Twenty-one people participated in the secondary survey I used to identify the subscales where I should focus my analysis. Respondents were asked to rank the top ten subscales in order of importance as related to the supervisor-subordinate relationship.

I analyzed the data with different methods. Leadership, Communication and Supervision received the strongest support with all three methods. First, I counted the number of blanks; a blank indicates the respondent doesn't consider the subscale important enough to be ranked. Next, I calculated the average response for the responses given. Finally, I calculated the average response if I assigned all the blanks a rank of zero and included the zeros in the summation. Results of the secondary survey are in Appendix F. These three subscales are comprised of twenty-nine items.

Table 3. Secondary Survey Results.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Number of Blanks</th>
<th>Average of Responses</th>
<th>Overall Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership</td>
<td>0</td>
<td>8.19</td>
<td>8.19</td>
</tr>
<tr>
<td>Communication</td>
<td>1</td>
<td>7.20</td>
<td>6.86</td>
</tr>
<tr>
<td>Supervision</td>
<td>1</td>
<td>5.89</td>
<td>5.95</td>
</tr>
</tbody>
</table>
Primary Survey

The remainder of this chapter is devoted to the analysis of responses from the primary survey.

Responses. Responses for the three subscales identified as most important to a supervisor-subordinate relationship are summarized in the Table 4. Mean subscale scores for Leadership, Communication, and Supervision for supervisors and subordinates are calculated for each category: pre-test subjects, pre-test/post-test subjects at post-test, and post-test only subjects. Standard deviations within subscale scores for each category are also provided.

<table>
<thead>
<tr>
<th>Table 4. Descriptive Statistics of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE-TEST SUBJECTS</td>
</tr>
<tr>
<td>Experimental Group</td>
</tr>
<tr>
<td>Supervisor</td>
</tr>
<tr>
<td>n 5</td>
</tr>
<tr>
<td>Subordinate</td>
</tr>
<tr>
<td>n 5</td>
</tr>
<tr>
<td>Control Group</td>
</tr>
<tr>
<td>Supervisor</td>
</tr>
<tr>
<td>n 4</td>
</tr>
<tr>
<td>Subordinate</td>
</tr>
<tr>
<td>n 4</td>
</tr>
<tr>
<td>POST-TEST SUBJECTS</td>
</tr>
<tr>
<td>Pre-test and Post-test</td>
</tr>
<tr>
<td>Experimental Group</td>
</tr>
<tr>
<td>Supervisor</td>
</tr>
<tr>
<td>n 5</td>
</tr>
<tr>
<td>Subordinate</td>
</tr>
<tr>
<td>n 5</td>
</tr>
<tr>
<td>Control Group</td>
</tr>
<tr>
<td>Supervisor</td>
</tr>
<tr>
<td>n 4</td>
</tr>
<tr>
<td>Subordinate</td>
</tr>
<tr>
<td>n 4</td>
</tr>
<tr>
<td>Post-test Only</td>
</tr>
<tr>
<td>Experimental Group</td>
</tr>
<tr>
<td>Supervisor</td>
</tr>
<tr>
<td>n 5</td>
</tr>
<tr>
<td>Subordinate</td>
</tr>
<tr>
<td>n 5</td>
</tr>
<tr>
<td>Control Group</td>
</tr>
<tr>
<td>Supervisor</td>
</tr>
<tr>
<td>n 2</td>
</tr>
<tr>
<td>Subordinate</td>
</tr>
<tr>
<td>n 2</td>
</tr>
</tbody>
</table>

Measurement Instrument Internal Consistency. Cronbach's Alpha measures of reliability for the three subscales receiving the highest rankings in the primary survey are provided in Table 5.

<table>
<thead>
<tr>
<th>Table 5. Cronbach's Alpha for Each Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Items</td>
</tr>
<tr>
<td>$\alpha$</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

47
Research Hypotheses Testing.

My five hypotheses derive directly from the research sub-questions. These hypotheses are explored primarily through independent t-tests ($\alpha = 2 \times 0.025$) on the subscale differences within supervisor-subordinate pairs. Only paired responses are considered for hypothesis testing. The experimental group started with 10 supervisors and 16 subordinates. Ten pairs participated in the post-test. The control group started with 10 supervisors and 9 subordinates. By the time the post-test survey was distributed, only six supervisor-subordinate pairs remained.

There Will Be a Significant Difference between Experimental and Control Group Differences for the Post-test. My first hypothesis looks at all paired subjects participating in the post-test; it's the crux of my study. I compare experimental and control group post-test Leadership, Communication and Supervision subscale differences for supervisor and subordinate subscale scores. The design of my test is illustrated in Table 6. Table 7 contains the descriptive statistics for the responses used in the t-test calculation. The t-values comparing the groups for these subscales are presented in Table 8. There are no significant differences.

Table 6. Design of the Post-test T-test.

<table>
<thead>
<tr>
<th>Experimental</th>
<th>Pre-test</th>
<th>POST-TEST Pre-test Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td>[1,10]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th>COMMUNICATION</th>
<th>SUPERVISION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n)</td>
<td>(\bar{x})</td>
<td>(s)</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.84</td>
<td>0.288</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.48</td>
<td>0.517</td>
</tr>
<tr>
<td>Post-test Only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.48</td>
<td>0.369</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.70</td>
<td>0.288</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>4</td>
<td>4.33</td>
<td>0.568</td>
</tr>
<tr>
<td>Subordinate</td>
<td>4</td>
<td>4.98</td>
<td>1.069</td>
</tr>
<tr>
<td>Post-test Only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>2</td>
<td>5.44</td>
<td>0.795</td>
</tr>
<tr>
<td>Subordinate</td>
<td>2</td>
<td>5.31</td>
<td>0.619</td>
</tr>
</tbody>
</table>
Table 8. Post-test T-test Results.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th>COMMUNICATION</th>
<th>SUPERVISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>(t_{\text{observed}})</td>
<td>1.2925</td>
<td>0.5035</td>
<td>0.2943</td>
</tr>
<tr>
<td>(t_{\text{critical}})</td>
<td>2.145</td>
<td>2.145</td>
<td>2.145</td>
</tr>
<tr>
<td>df</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

*denotes a significant difference between the experimental and control group responses at \(\alpha = 0.05\).

There Will Be No Significant Difference between the Experimental and Control Group Differences for the Pre-test. My second hypothesis looks for differences between the two groups at the start of the study. I compare experimental and control group pre-test Leadership, Communication, and Supervision subscale differences for supervisor and subordinate subscale scores. The design of my test is illustrated in Table 9. Table 10 contains the descriptive statistics for the responses used in the t-test calculation. The t-values comparing the groups for these subscales are presented in Table 11. There are no significant differences.

Table 9. Design of the Pre-test T-test.

<table>
<thead>
<tr>
<th></th>
<th>PRE-TEST</th>
<th>POST-TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>[1, 5]</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>[a, d]</td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Descriptive Statistics for the Pre-test T-test Responses.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>(\bar{x})</td>
<td>s</td>
<td>n</td>
<td>(\bar{x})</td>
<td>s</td>
<td>n</td>
<td>(\bar{x})</td>
<td>s</td>
</tr>
<tr>
<td>Experimental Group</td>
<td>5</td>
<td>4.66</td>
<td>0.385</td>
<td>5</td>
<td>5.14</td>
<td>0.702</td>
<td>5</td>
<td>4.84</td>
<td>0.513</td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>5.24</td>
<td>0.764</td>
<td>5</td>
<td>5.48</td>
<td>0.841</td>
<td>5</td>
<td>5.52</td>
<td>0.854</td>
</tr>
<tr>
<td>Subordinate</td>
<td>4</td>
<td>4.25</td>
<td>0.465</td>
<td>4</td>
<td>4.75</td>
<td>0.557</td>
<td>4</td>
<td>4.70</td>
<td>0.408</td>
</tr>
<tr>
<td>Control Group</td>
<td>4</td>
<td>4.90</td>
<td>0.688</td>
<td>4</td>
<td>5.00</td>
<td>1.414</td>
<td>4</td>
<td>4.98</td>
<td>0.741</td>
</tr>
</tbody>
</table>

Table 11. Pre-test T-test Results.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th>COMMUNICATION</th>
<th>SUPERVISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>(t_{\text{observed}})</td>
<td>0.8484</td>
<td>1.1145</td>
<td>0.9949</td>
</tr>
<tr>
<td>(t_{\text{critical}})</td>
<td>2.365</td>
<td>2.365</td>
<td>2.365</td>
</tr>
<tr>
<td>df</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

*denotes a significant difference between the experimental and control group responses at \(\alpha = 0.05\).
The Instrument Will Have an Effect on the Subjects. In my third hypothesis I'm looking for some indication the instrument had some influence on the subjects. Responding to the instrument might increase awareness; supervisors and subordinates could become more aware of supervisor task performance. To test this hypothesis, all paired subjects participating in the post-test are considered. This time, the paired subjects are divided according to whether they participated in the pre-test. I compare pre-test and post-test at post-test and post-test only for Leadership, Communication and Supervision subscale differences for supervisor and subordinate subscale scores. The design of my test is illustrated in Table 12. Table 13 contains the descriptive statistics for the responses used in the t-test calculation. The t-values comparing the groups for these subscales are presented in Table 14. There are no significant differences.

Table 12. Design of the Pre-test and Post-test versus Post-test Only Post-test T-test.

<table>
<thead>
<tr>
<th></th>
<th>PRE-TEST</th>
<th>POST-TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>Experimental</td>
<td>[1, 5]</td>
<td>[6, 10]</td>
</tr>
<tr>
<td>Control</td>
<td>[a, d]</td>
<td>[e, f]</td>
</tr>
</tbody>
</table>

Table 13. Descriptive Statistics for the Pre-test and Post-test versus Post-test Only T-test Responses.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th>COMMUNICATION</th>
<th>SUPERVISION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>x</td>
<td>s</td>
</tr>
<tr>
<td>Pre-test and Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.84</td>
<td>0.288</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.48</td>
<td>0.517</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>4</td>
<td>4.33</td>
<td>0.568</td>
</tr>
<tr>
<td>Subordinate</td>
<td>4</td>
<td>4.98</td>
<td>1.069</td>
</tr>
<tr>
<td>Post-test Only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.48</td>
<td>0.369</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.70</td>
<td>0.288</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>2</td>
<td>5.44</td>
<td>0.795</td>
</tr>
<tr>
<td>Subordinate</td>
<td>2</td>
<td>5.31</td>
<td>0.619</td>
</tr>
</tbody>
</table>
Table 14. Pre-test and Post-test versus Post-test Only Post-test T-test Results.

<table>
<thead>
<tr>
<th>α</th>
<th>LEADERSHIP</th>
<th>COMMUNICATION</th>
<th>SUPERVISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( t_{\text{observed}} )</td>
<td>0.4229</td>
<td>0.0292</td>
<td>0.4727</td>
</tr>
<tr>
<td>( t_{\text{critical}} )</td>
<td>2.145</td>
<td>2.145</td>
<td>2.145</td>
</tr>
<tr>
<td>df</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

* denotes a significant difference between the pre-test and post-test and post-test only responses at \( \alpha = 0.05 \).

**Differences between the Supervisor and Paired Subordinate Are Affected by the Treatment.** My fourth hypothesis examines whether the treatment had an impact on the change in differences between the pre-test and post-test. To look at change, I concentrate on pairs participating in the pre-test. To understand what this test is doing, think of dependent t-tests. Rather than performing two dependent t-tests, the differences within the pair comparisons are computed; the average of these differences for the two groups are compared. Because the differences between the post-test and pre-test subscale scores for each pair of differences are computed before the t-test analysis, a dependent t-test is inappropriate despite the fact that only the pairs completing both the pre-test and post-test are considered. I compare experimental group and control group longitudinal changes in Leadership, Communication, and Supervision subscale differences for supervisor and subordinate subscale scores. The design of my test is illustrated in Table 15. Table 16 contains the descriptive statistics for the responses used in the t-test calculation. The t-values comparing the groups for these subscales are presented in Table 17. There is a significant difference in Leadership, but the interpretation of this change isn't straightforward. After training, the difference between the experimental group supervisors and subordinates increased, with both giving the supervisor more favorable ratings; subordinates' supervisor ratings increased slightly more than self-evaluations. In the control group, both supervisors and subordinates gave more favorable ratings in the post-test, but the changes were more consistent with each other.

Table 15. Design of the Pre-test Post-test Differences T-test.

<table>
<thead>
<tr>
<th>Experimental</th>
<th>PRE-TEST</th>
<th>POST-TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Pre-test</td>
<td>[1, 5]</td>
</tr>
<tr>
<td></td>
<td>difference</td>
<td>difference</td>
</tr>
<tr>
<td></td>
<td>[a, d]</td>
<td>[a, d]</td>
</tr>
</tbody>
</table>

51
Table 16. Descriptive Statistics for the Pre-test Post-test Differences T-test Responses.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th></th>
<th>COMMUNICATION</th>
<th></th>
<th>SUPERVISION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>x</td>
<td>s</td>
<td>n</td>
<td>x</td>
<td>s</td>
</tr>
<tr>
<td><strong>Experimental Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.66</td>
<td>0.385</td>
<td>5</td>
<td>5.14</td>
<td>0.702</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.24</td>
<td>0.764</td>
<td>5</td>
<td>5.48</td>
<td>0.841</td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.84</td>
<td>0.288</td>
<td>5</td>
<td>5.08</td>
<td>0.536</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.48</td>
<td>0.517</td>
<td>5</td>
<td>5.66</td>
<td>0.760</td>
</tr>
<tr>
<td><strong>Control Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>4</td>
<td>4.25</td>
<td>0.465</td>
<td>4</td>
<td>4.75</td>
<td>0.557</td>
</tr>
<tr>
<td>Subordinate</td>
<td>4</td>
<td>4.90</td>
<td>0.688</td>
<td>4</td>
<td>5.00</td>
<td>1.414</td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>4</td>
<td>4.33</td>
<td>0.568</td>
<td>4</td>
<td>4.58</td>
<td>0.562</td>
</tr>
<tr>
<td>Subordinate</td>
<td>4</td>
<td>4.98</td>
<td>1.669</td>
<td>4</td>
<td>4.90</td>
<td>1.764</td>
</tr>
</tbody>
</table>

Table 17. Pre-test Post-test Differences T-test Results.

<table>
<thead>
<tr>
<th>α</th>
<th>LEADERSHIP</th>
<th>COMMUNICATION</th>
<th>SUPERVISION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t_{observed}</td>
<td>t_{critical}</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>2.941*</td>
<td>2.365</td>
<td>2.365</td>
</tr>
<tr>
<td></td>
<td>1.213</td>
<td>2.365</td>
<td>2.365</td>
</tr>
<tr>
<td>0.05</td>
<td>1.400</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

* denotes a significant difference between the experimental and control group pre-test supervisor-subordinate differences and post-test supervisor and subordinate differences for the same pairs of supervisors and subordinates at $\alpha = 0.05$.

**Experimental Group Subjects Who Provide Responses for the First Time Will Differ between Pre- and Post-tests.** My fifth hypothesis is another way to examine the impact of the treatment. Except for the treatment, operations within the experimental sub-organization continued without experimental influence between the pre-test and the post-test. Supervisor-subordinate relationships, established prior to the start of the study, were maintained for the duration of the study. Experimental group subjects were randomly assigned pre-tests. Any differences between the subject the first time they completed a survey can be attributed to the treatment. I compare experimental group pre-test and experimental group post-test only Leadership, Supervision, and Communication subscale differences for supervisor and subordinate subscale scores. The design of my test is illustrated in Table 18. Table 19 contains the descriptive statistics for the responses used in the t-test calculation. The t-values comparing the groups for these subscales are presented in Table 20. There are no significant differences.
Table 18. Design of the First Time Experimental Group T-test.

<table>
<thead>
<tr>
<th>Experimental Control</th>
<th>PRE-TEST</th>
<th>POST-TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Pre-test</td>
</tr>
<tr>
<td></td>
<td>[1, 5]</td>
<td>[6, 10]</td>
</tr>
</tbody>
</table>

Table 19. Descriptive Statistics for the First Time Experimental T-test Responses.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th></th>
<th></th>
<th>COMMUNICATION</th>
<th></th>
<th></th>
<th>SUPERVISION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>(\bar{x})</td>
<td>s</td>
<td>n</td>
<td>(\bar{x})</td>
<td>s</td>
<td>n</td>
<td>(\bar{x})</td>
</tr>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.66</td>
<td>0.385</td>
<td>5</td>
<td>5.14</td>
<td>0.702</td>
<td>5</td>
<td>4.84</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.24</td>
<td>0.764</td>
<td>5</td>
<td>5.48</td>
<td>0.841</td>
<td>5</td>
<td>5.52</td>
</tr>
<tr>
<td>Post-test Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.48</td>
<td>0.369</td>
<td>5</td>
<td>4.60</td>
<td>0.985</td>
<td>5</td>
<td>4.60</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>5.70</td>
<td>0.288</td>
<td>5</td>
<td>5.47</td>
<td>0.769</td>
<td>5</td>
<td>5.38</td>
</tr>
</tbody>
</table>

Table 20. First Time Experimental Group Participants T-test Results.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th></th>
<th></th>
<th>COMMUNICATION</th>
<th></th>
<th></th>
<th>SUPERVISION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t_{observed})</td>
<td>0.795</td>
<td>1.879</td>
<td>0.4824</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t_{critical})</td>
<td>2.306</td>
<td>2.306</td>
<td>2.306</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\* denotes a significant difference between the differences in paired scores for first time test responses at \(\alpha = 0.05\).
NOTES

1. All subject pairs provided sufficient usable responses for the three subscale calculations, allowing the greatest possible degrees of freedom for the sample.

2. Only items containing only usable responses are used in the calculation; the number of items for each subscale meeting this requirement are the number of items used in each calculation: Leadership: #83, #86; Communication: #69, #71; and Supervision: #22, #23, #24, #34. All subjects providing post-test responses are used, including unpaired subjects. The responses of a part-time student (not part of the population) are ignored.

3. The formula used to calculate the $t_{observed}$ is for unknown variances assumed equal.

4. I considered there might be a preferred direction indicated by a positive or negative subscale difference. If there were a preferred direction, more significant differences would have been observed (with a one-sided $t$-test). Since the responses aren't consistent with previous studies, I couldn't make this judgment.

5. I'm interested in the magnitudes of the differences between paired supervisor and subordinate perspectives. Throughout the calculations presented in Chapter 4, I use the absolute differences. Magnitudes, not directions, are analyzed.

6. To arrive at this collection of responses, subjects completing only a pre-test, a student working part-time, both a control group supervisor and the paired subordinate who submitted a blank opscan form, and three experimental group subordinates are eliminated. I'm interested in pairs; where a supervisor has two subordinates, one has to be ignored. Two of these subordinates were selected because there were blanks on their opscan forms; the third was selected by the toss of a coin.
Chapter 5
EXPLORATORY ANALYSIS

Chapter 5 contains a more comprehensive, less detailed analysis of the differences than Chapter 4 and some observations on the responses.

Comprehensive Analysis

When I began my study, I had two erroneous expectations: 1.) supervisors would evaluate themselves more favorably than subordinates rated them and 2.) the fourteen most important tasks for first-line supervisors (U.S. Merit Systems Protection Board, 1992) would have significant differences between the experimental and control groups. My instrument is broad. Following the 118 task-oriented questions, there are items related to general impressions. I explore all these avenues.

Supervisors Don't Always Rate Themselves More Favorably Than Their Subordinates Rate Them. Based on the literature review, I expected to find supervisors rating themselves more favorably than their subordinates rate them. I was wrong. To get a picture of the relationship between paired responses, I examined plots of the mean subscale scores for supervisors and subordinates for each of the two groups (experimental and control). Figures 4 through 9, pre-test subjects, pre-test and post-test subjects at post-test, and post-test only subjects, include mean subscale scores for the subjects used for my primary analysis. For ease of comparison, the supervisors and subordinates for common categories are presented on the same axis.

Overwhelmingly, except in Figure 9, subordinates gave their supervisors more favorable evaluations than supervisors gave themselves in all of the experimental group categories and the control group pre-test. Both control group post-test category responses tended to have supervisors rating themselves higher than their subordinates rate them; in most cases, the pre-test and post-test subordinates reduced their pre-test scores. This big-picture view is consistent with my primary analysis. In a comparison of the mean subscale scores, Table 4, only control group post-test only supervisors in the subscales for leadership and communication rated themselves more favorably than their subordinates rated them. A comparison of the median subscale scores for the same three subscales, Table 21, is similar. In this view, the control group post-test only supervisors rate themselves more favorably for the leadership subscale.
Figure 4. Mean Pre-test Subscale Scores for Experimental Group Subjects.

Figure 5. Mean Pre-test Subscale Scores for Control Group Subjects.
Figure 6. Mean Post-test Subscale Scores for Experimental Group Pre-test and Post-test Subjects.

Figure 7. Mean Post-test Subscale Scores for Control Group Pre-test and Post-test Subjects.
Figure 8. Mean Post-test Subscale Scores for Experimental Group Post-test Only Subjects.

Figure 9. Mean Post-test Subscale Scores for Control Group Post-test Only Subjects.
Table 21. Descriptive Statistics of Responses.

<table>
<thead>
<tr>
<th></th>
<th>LEADERSHIP</th>
<th>COMMUNICATION</th>
<th>SUPERVISION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Median</td>
<td>n</td>
</tr>
<tr>
<td>PRE-TEST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>6.0</td>
<td>5</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>4</td>
<td>5.0</td>
<td>4</td>
</tr>
<tr>
<td>Subordinate</td>
<td>4</td>
<td>5.0</td>
<td>4</td>
</tr>
<tr>
<td>POST-TEST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test and Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>6.0</td>
<td>5</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>4</td>
<td>5.0</td>
<td>4</td>
</tr>
<tr>
<td>Subordinate</td>
<td>4</td>
<td>5.0</td>
<td>4</td>
</tr>
<tr>
<td>Post-test Only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>5</td>
<td>4.0</td>
<td>5</td>
</tr>
<tr>
<td>Subordinate</td>
<td>5</td>
<td>6.0</td>
<td>5</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>2</td>
<td>6.0</td>
<td>2</td>
</tr>
<tr>
<td>Subordinate</td>
<td>2</td>
<td>5.0</td>
<td>2</td>
</tr>
</tbody>
</table>

A look at the median subscale scores in Figures 10 through 15 provides another view of the same general trend. Keeping the subscale level of analysis and using the same subscale scores calculated for hypothesis testing, I examine the measured directions of the supervisor and subordinate subscale score differences. Again subordinate evaluations apparently are often more favorable than self evaluations. For the experimental group, this trend is fairly consistent. For the control group, higher supervisor median subscale scores can be seen in the pre-test and post-test responses, particularly in the post-test. I use these median subscale scores for nonparametric statistical analysis.
Figure 10. Median Pre-test Subscale Scores for Experimental Group Subjects.

Figure 11. Median Pre-test Subscale Scores for Control Group Subjects.
Figure 12. Median Post-test Subscale Scores for Experimental Group Pre-test and Post-test Subjects.

Figure 13. Median Post-test Subscale Scores for Control Group Pre-test and Post-test Subjects.
Figure 14. Median Post-test Subscale Scores for Experimental Group Post-test Only Subjects.

Figure 15. Median Post-test Subscale Scores for Control Group Post-test Only Subjects.
Since the experimental group subordinates tended to have higher subscale scores than their matched supervisor's scores, my calculations for nonparametric statistical analysis are based on subordinate minus supervisor scores for each subscale. The differences between the supervisors and their matched subordinates are ranked; the ranks for each subscale are summed. Tables 22 and 24 present the results of the Wilcoxon-Mann-Whitney tests. Tables 23 and 25 contain descriptive statistics for the subscales with significant differences between the experimental and control groups; the mean and variance reflect magnitude while the median incorporates direction.

Table 22. Pre-test Subscale Ranking Results.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Experimental Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number of cases</td>
<td>Σ ranks</td>
</tr>
<tr>
<td>W. U. Planning</td>
<td>0.5714</td>
<td>4 16</td>
</tr>
<tr>
<td>W. U. Guidance</td>
<td>0.6071</td>
<td>5 22</td>
</tr>
<tr>
<td>Budgeting</td>
<td>0.5000</td>
<td>3 11</td>
</tr>
<tr>
<td>Mat'l Res. Admin.</td>
<td>n.e.v.d.</td>
<td>3 0</td>
</tr>
<tr>
<td>Supervision</td>
<td>0.4524</td>
<td>5 26</td>
</tr>
<tr>
<td>Personnel Mgmt.</td>
<td>0.3500</td>
<td>3 12</td>
</tr>
<tr>
<td>External Awareness</td>
<td>0.3429</td>
<td>4 20</td>
</tr>
<tr>
<td>Interpretation</td>
<td>0.2429</td>
<td>4 21</td>
</tr>
<tr>
<td>Representation</td>
<td>0.0143*</td>
<td>4 26</td>
</tr>
<tr>
<td>Coordination</td>
<td>0.0429*</td>
<td>4 24.5</td>
</tr>
<tr>
<td>W. U. Monitoring</td>
<td>0.5571</td>
<td>4 18</td>
</tr>
<tr>
<td>Program Evaluation</td>
<td>0.9429</td>
<td>3 12</td>
</tr>
<tr>
<td>Communication</td>
<td>0.5913</td>
<td>5 24.5</td>
</tr>
<tr>
<td>Interpersonal Sens.</td>
<td>0.2063</td>
<td>5 29</td>
</tr>
<tr>
<td>Leadership</td>
<td>0.6349</td>
<td>5 24</td>
</tr>
<tr>
<td>Flexibility</td>
<td>0.4088</td>
<td>5 26.5</td>
</tr>
<tr>
<td>Action Orientation</td>
<td>0.3143</td>
<td>4 18</td>
</tr>
<tr>
<td>Results Focus</td>
<td>0.5913</td>
<td>5 24.5</td>
</tr>
<tr>
<td>Broad Perspectives</td>
<td>0.2429</td>
<td>4 21</td>
</tr>
<tr>
<td>Strategic View</td>
<td>0.4524</td>
<td>5 26</td>
</tr>
<tr>
<td>Environmental Sens</td>
<td>0.0786</td>
<td>4 23.5</td>
</tr>
</tbody>
</table>

* denotes a significant difference between experimental and control group responses at α = 0.05.

n.e.v.d. means there isn't enough usable data to perform the calculation.

Table 23. Descriptive Statistics of Pre-test Paired Difference Responses with Significant Differences at α = 0.05.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>n</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>¥</td>
<td>¥</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¥</td>
<td>¥</td>
</tr>
<tr>
<td>Experimental</td>
<td>4</td>
<td>1.08</td>
<td>0.50</td>
</tr>
<tr>
<td>Control</td>
<td>4</td>
<td>0.46</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Table 24. Post-test Subscale Ranking Results.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Experimental Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p value</td>
<td>number of cases</td>
</tr>
<tr>
<td>1. W. U. Planning</td>
<td>0.3516</td>
<td>8</td>
</tr>
<tr>
<td>2. W. U. Guidance</td>
<td>0.1864</td>
<td>9</td>
</tr>
<tr>
<td>3. Budgeting</td>
<td>n.e.v.d.</td>
<td>6</td>
</tr>
<tr>
<td>4. Mat'l Res. Admin.</td>
<td>n.e.v.d.</td>
<td>3</td>
</tr>
<tr>
<td>5. Supervision</td>
<td>0.1317</td>
<td>10</td>
</tr>
<tr>
<td>6. Personnel Mgmt.</td>
<td>0.4040</td>
<td>8</td>
</tr>
<tr>
<td>7. External Awareness</td>
<td>0.0415*</td>
<td>9</td>
</tr>
<tr>
<td>8. Interpretation</td>
<td>0.0183*</td>
<td>10</td>
</tr>
<tr>
<td>9. Representation</td>
<td>0.0567</td>
<td>9</td>
</tr>
<tr>
<td>10. Coordination</td>
<td>0.0332*</td>
<td>9</td>
</tr>
<tr>
<td>11. W. U. Monitoring</td>
<td>0.0376*</td>
<td>10</td>
</tr>
<tr>
<td>12. Program Evaluation</td>
<td>0.1489</td>
<td>9</td>
</tr>
<tr>
<td>13. Communication</td>
<td>0.2811</td>
<td>10</td>
</tr>
<tr>
<td>14. Interpersonal Sens.</td>
<td>0.0663</td>
<td>10</td>
</tr>
<tr>
<td>15. Leadership</td>
<td>0.0210*</td>
<td>10</td>
</tr>
<tr>
<td>16. Flexibility</td>
<td>0.0156*</td>
<td>10</td>
</tr>
<tr>
<td>17. Action Orientation</td>
<td>0.0663</td>
<td>10</td>
</tr>
<tr>
<td>18. Results Focus</td>
<td>0.1442</td>
<td>10</td>
</tr>
<tr>
<td>19. Broad Perspectives</td>
<td>0.1961</td>
<td>9</td>
</tr>
<tr>
<td>20. Strategic View</td>
<td>0.0820</td>
<td>10</td>
</tr>
<tr>
<td>21. Environmental Sens.</td>
<td>0.0814</td>
<td>9</td>
</tr>
</tbody>
</table>

* denotes a significant difference between experimental and control group responses at α = 0.05.

n.e.v.d means there isn't enough usable data to perform the calculation.

Table 25 Descriptive Statistics of Post-test Paired Response Differences with Significant Differences at α = 0.05.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>7</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>x</td>
<td>s</td>
</tr>
<tr>
<td>EXPERIMENTAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test/Post-test</td>
<td>4</td>
<td>1.06</td>
<td>0.75</td>
</tr>
<tr>
<td>Post-test Only</td>
<td>5</td>
<td>0.48</td>
<td>0.37</td>
</tr>
<tr>
<td>CONTROL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test/Post-test</td>
<td>3</td>
<td>1.42</td>
<td>0.72</td>
</tr>
<tr>
<td>Post-test Only</td>
<td>2</td>
<td>0.25</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Subscale                  | 11 | 15 | 19 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>x</td>
<td>s</td>
</tr>
<tr>
<td>EXPERIMENTAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test/Post-test</td>
<td>5</td>
<td>0.67</td>
<td>0.71</td>
</tr>
<tr>
<td>Post-test Only</td>
<td>5</td>
<td>0.90</td>
<td>0.70</td>
</tr>
<tr>
<td>CONTROL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test/Post-test</td>
<td>3</td>
<td>1.60</td>
<td>1.06</td>
</tr>
<tr>
<td>Post-test Only</td>
<td>2</td>
<td>0.60</td>
<td>0.28</td>
</tr>
</tbody>
</table>

64
Significant Differences. Results of the t-tests for all twenty-one subscales are presented in Table 26. I’ll briefly review the reasons for each test and list the significant differences. The original group sizes are small. Unusable data further reduces the differences included in many of the calculations. I use magnitude, not direction, in my calculations. A list of the subscales is in Appendix G, or the subscales and their descriptive items can be examined in Appendix A, my original instrument.

Post-test: I’m trying to see if the treatment affected the experimental group enough to observe a difference from the control group using a post-test t-test. This is the crux of my study. I want to see where the two groups differ after the experimental group receives training. The subscale for work unit guidance is significantly different between the two groups. Six items comprise this subscale.

Pre-test: I’m trying to establish a baseline for comparison between the experimental and control groups with the pre-test t-test. Since I didn’t have random assignment, this step is very important. I know the two groups aren’t statistically equivalent, so I want to see where they differ. None of the subscales have statistically significant differences between the experimental and control groups. I hoped to find no significant differences between the experimental and control groups at the start of my study.

P&P vs PO (Pre-test and Post-test versus Post-test Only): I’m trying to determine whether the instrument has an effect when I compare the post-test responses of subjects who took both the pre-test and the post-test with subjects taking only the post-test. There are two significant differences between the experimental and control groups. The differences for the work unit planning and external awareness subscales are significant. Four items comprise each of these subscales.

Differences: I’m trying to see whether the treatment affected the changes in paired data when I compare the absolute values of the differences between pre-test differences and post-test differences for the experimental and control groups. There are five significant differences between the experimental and control groups. These differences are in the subscales for work unit guidance with six items, program evaluation with five items, leadership with eight items, broad perspectives focus with four items, and strategic view with six items.
1st Time: I'm trying to see whether the impact of the treatment can be observed in the experimental group when I compare the pre-test responses and the post-test only responses. Since the subjects in the experimental group were randomly assigned to participate in the pre-test, both halves of the experimental group should have been as statistically equivalent as they could be. Any differences in the first-time responses must be due to the treatment. Only the subscale for work unit planning, comprised of four items, differs significantly between experimental group subjects participating the first time at the pre-test and subjects participating the first time at the post-test.

Table 26. Subscale T-test Results.

<table>
<thead>
<tr>
<th>subscale</th>
<th>Post-test</th>
<th>Pre-test</th>
<th>P &amp; P vs PO</th>
<th>Differences</th>
<th>1st Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df  t-test</td>
<td>df  t-test</td>
<td>df  t-test</td>
<td>df  t-test</td>
<td>df  t-test</td>
</tr>
<tr>
<td>1</td>
<td>9  0.12</td>
<td>5  0.89</td>
<td>9  2.78*</td>
<td>2  1.00</td>
<td>6  2.80*</td>
</tr>
<tr>
<td>2</td>
<td>10 2.29*</td>
<td>6  0.80</td>
<td>10 2.03</td>
<td>5  4.20*</td>
<td>7  2.00</td>
</tr>
<tr>
<td>3</td>
<td>6  0.63</td>
<td>4  0.72</td>
<td>6  0.76</td>
<td>2  0.60</td>
<td>4  0.30</td>
</tr>
<tr>
<td>4</td>
<td>n.e.v.d.</td>
<td>n.e.v.d.</td>
<td>n.e.v.d.</td>
<td>n.e.v.d.</td>
<td>3  1.48</td>
</tr>
<tr>
<td>5</td>
<td>14 0.29</td>
<td>7  0.99</td>
<td>14 0.47</td>
<td>7  1.40</td>
<td>8  0.48</td>
</tr>
<tr>
<td>6</td>
<td>10 0.16</td>
<td>4  0.13</td>
<td>10 0.72</td>
<td>3  0.06</td>
<td>6  0.41</td>
</tr>
<tr>
<td>7</td>
<td>12 0.52</td>
<td>6  0.48</td>
<td>12 2.62*</td>
<td>5  0.15</td>
<td>7  1.51</td>
</tr>
<tr>
<td>8</td>
<td>14 0.74</td>
<td>6  2.07</td>
<td>14 0.36</td>
<td>6  2.22</td>
<td>7  0.08</td>
</tr>
<tr>
<td>9</td>
<td>13 0.11</td>
<td>6  1.71</td>
<td>13 0.70</td>
<td>6  1.01</td>
<td>7  1.08</td>
</tr>
<tr>
<td>10</td>
<td>13 1.18</td>
<td>6  2.12</td>
<td>13 0.34</td>
<td>6  0.40</td>
<td>7  0.70</td>
</tr>
<tr>
<td>11</td>
<td>13 0.98</td>
<td>6  0.95</td>
<td>13 0.52</td>
<td>5  1.53</td>
<td>7  0.79</td>
</tr>
<tr>
<td>12</td>
<td>12 0.24</td>
<td>5  0.47</td>
<td>12 0.35</td>
<td>4  3.86*</td>
<td>6  0.53</td>
</tr>
<tr>
<td>13</td>
<td>14 0.50</td>
<td>7  1.11</td>
<td>14 0.03</td>
<td>7  1.22</td>
<td>8  1.88</td>
</tr>
<tr>
<td>14</td>
<td>14 0.39</td>
<td>7  0.66</td>
<td>14 0.37</td>
<td>7  0.98</td>
<td>8  0.62</td>
</tr>
<tr>
<td>15</td>
<td>14 1.29</td>
<td>7  0.85</td>
<td>14 0.42</td>
<td>7  2.94*</td>
<td>8  0.80</td>
</tr>
<tr>
<td>16</td>
<td>14 0.03</td>
<td>7  0.27</td>
<td>14 0.64</td>
<td>7  0.87</td>
<td>8  0.37</td>
</tr>
<tr>
<td>17</td>
<td>14 1.23</td>
<td>5  0.40</td>
<td>14 1.73</td>
<td>5  1.08</td>
<td>7  0.13</td>
</tr>
<tr>
<td>18</td>
<td>14 0.68</td>
<td>7  0.45</td>
<td>14 1.28</td>
<td>7  1.17</td>
<td>8  0.42</td>
</tr>
<tr>
<td>19</td>
<td>13 0.29</td>
<td>6  0.58</td>
<td>13 0.85</td>
<td>5  3.83*</td>
<td>7  1.05</td>
</tr>
<tr>
<td>20</td>
<td>14 0.11</td>
<td>7  0.54</td>
<td>14 0.62</td>
<td>7  5.10*</td>
<td>8  0.81</td>
</tr>
<tr>
<td>21</td>
<td>13 0.81</td>
<td>6  0.22</td>
<td>13 0.26</td>
<td>6  1.99</td>
<td>7  0.27</td>
</tr>
</tbody>
</table>

P & P vs PO means Pre-test and Post-test versus Post-test Only
n.e.v.d. means there is not enough usable data to perform the calculation.
* denotes a significant difference at α = 0.05.
Within Pair Changes: the pairs participating in the pre-test and the post-test provide an opportunity to examine the data a little differently. With these subjects, I can look for differences between the pre-test and the post-test. The dependent t-test results for the experimental and control group are presented in Table 27. For the experimental group, the subscales for work unit monitoring and program evaluation, each composed of five items, are significantly different. For the control group, there are no significant differences.

Table 27. Dependent T-test Results.

<table>
<thead>
<tr>
<th>EXPERIMENTAL GROUP</th>
<th>SUBSCALE NUMBER</th>
<th>CONTROL GROUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>t-test</td>
<td>df</td>
</tr>
<tr>
<td>2</td>
<td>0.400</td>
<td>n.e.v.d.</td>
</tr>
<tr>
<td>4</td>
<td>0.3203</td>
<td>3.2857</td>
</tr>
<tr>
<td>1</td>
<td>0.7143</td>
<td>n.e.v.d.</td>
</tr>
<tr>
<td>4</td>
<td>1.8708</td>
<td>0.6234</td>
</tr>
<tr>
<td>2</td>
<td>0.1358</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.2538</td>
<td>1.4626</td>
</tr>
<tr>
<td>3</td>
<td>0.2966</td>
<td>1.2026</td>
</tr>
<tr>
<td>3</td>
<td>1.0874</td>
<td>0.7295</td>
</tr>
<tr>
<td>3</td>
<td>1.6330</td>
<td>1.0604</td>
</tr>
<tr>
<td>3</td>
<td>4.1575*</td>
<td>1.6938</td>
</tr>
<tr>
<td>2</td>
<td>5.0000*</td>
<td>0.5470</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>0.4484</td>
</tr>
<tr>
<td>4</td>
<td>0.6224</td>
<td>0.3730</td>
</tr>
<tr>
<td>4</td>
<td>1.4505</td>
<td>0.3721</td>
</tr>
<tr>
<td>4</td>
<td>0.6271</td>
<td>0.3712</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>0.3172</td>
</tr>
<tr>
<td>4</td>
<td>0.5599</td>
<td>0.1703</td>
</tr>
<tr>
<td>3</td>
<td>0.9608</td>
<td>0.2524</td>
</tr>
<tr>
<td>4</td>
<td>0.2453</td>
<td>0.3593</td>
</tr>
<tr>
<td>3</td>
<td>1.3207</td>
<td>0.1750</td>
</tr>
</tbody>
</table>

n.e.v.d. means there is not enough usable data to perform the calculation.

* denoted a significant difference at $\alpha = 0.05$. 
Fourteen Most Important Tasks. I looked at the fourteen items addressing the tasks identified by MSPB as important to the success of a supervisor. (See Appendix B.) The graph in Figure 16 illustrates the absolute differences between paired supervisor and subordinate responses for experimental and control groups pre- and post-tests. No trends or patterns are obvious.

Figure 16. Absolute Differences between Paired Supervisor and Subordinate Median Scores for the Fourteen Most Important Tasks.
General Impressions. There are seven instrument items related to general impressions: efficient use of time, quality of work, multiple demand performance, creativity, skill knowledge, self contributions to the organization, and overall quality/effectiveness. The graph in Figure 17 illustrates the absolute differences between paired supervisor and subordinate responses for experimental and control groups pre- and post-tests. No trends or patterns are obvious.

![Graph](image)

Figure 17. Absolute Differences between Paired Supervisor and Subordinate Median Scores for General Impression Items.

Some Observations
When I began to look at the data, I was surprised to see so many instances of supervisors being unable to rate their own performance on a given task. I also didn't expect to see many cases of supervisors saying they don't perform a given task, after all, the instrument is based on years of research. I supposed that subordinates unable to rate their supervisor on a given task in the pre-test would have more of an opinion on the post-test. And I certainly didn't expect subordinates to give exceptional ratings to supervisors who cannot judge their own performance or say that task is not performed. In short, there are many unexpected turns in the data. I concentrated on pre-test and post-test pairs when I pursued these tantalizing leads.

I can't find any consistency between pre-test and post-test responses. The same task that receives a doesn't do in the pre-test can receive an exceptional rating in the post-test. The opposite also occurs. A weakness from the supervisor's perspective can be exceptional from the subordinate's
perspective. The opposite also occurred. The twenty-two items receiving no "doesn't do this task" or "I can't judge" ratings are listed in Appendix H.

The Doesn't Do This Task Response. In only 3 of 30 pre-test cases of the subordinate saying the supervisor doesn't do a task does the supervisor agree. While all the control group subordinates gave this post-test response at least once, two of five experimental group subordinates gave similar responses. There were 7 instances where supervisors gave themselves exceptional post-test ratings at the same time their subordinates said the supervisors don't do that task. Of the 112 post-test instances of subordinates saying a task isn't done, 32 gave exceptional ratings in the pre-test.

The I Can't Judge Response. Of the 74 cases where subordinates couldn't judge their supervisors on a given task in the pre-test, 19 gave exceptional ratings in the post-test. Of the 18 cases where supervisors couldn't judge their performance in the pre-test, 7 subordinates gave exceptional post-test ratings.

The Can Do This TaskExceptionally Well Response. Of the 127 cases where supervisors gave themselves exceptional post-test ratings, subordinates gave 81 exceptional; 20 above average; 13 acceptable; 1 barely acceptable; 3 not acceptable; 8 doesn't do; and 1 can't judge. These same supervisors gave some different responses in the pre-test: 51 above average; 11 acceptable; 2 don't do; and 1 can't judge. The three tasks with the most instances of supervisors giving themselves exceptional ratings are part of the supervision subscale (giving praise for good work, allowing employees to work without unnecessarily close supervision and doing journeyman work in an emergency).
Chapter 6
DISCUSSION

Chapter 6 contains an analysis of my results, improvements I would make if I were to do this study again, recommendations for managers considering TQM training and training evaluation, and suggested topics for future research.

Interpretation of My Results
While the focus of my analysis was three of twenty-one subscales from my measurement instrument, none were overlooked. In addition to presenting the results of the t-tests I performed for the hypothesis testing, I scrutinized the supervisor-subordinate pairs participating in both the pre- and post-tests. A comparison of the direction of evaluation differences was surprising. I had anticipated from the start of the study that supervisors would evaluate themselves more favorably than their paired subordinates would evaluate them. I didn't expect to see subordinates giving the higher ratings.

The frequency of unusable responses necessitated the calculation of subscale scores. In hindsight, I would choose a subscale approach over an individual item approach right from the start. The subscale provides more dimensions of a particular characteristic. A more complete picture of each effectiveness characteristic or management function is used for comparison, probably more closely mirroring the intent of the OPM originators of the Management Excellence Inventory, the precursor to the MSPB questionnaire that my instrument is based on.

I'm interested in the magnitude of the differences between supervisor and subordinate perceptions. Except in my nonparametric statistical analysis, I use the absolute differences between supervisor and subordinate subscale score so magnitudes of differences, not directions of differences, are analyzed. I cannot say whether it's better for the direction of supervisor subscale score subtracted from the subordinate subscale score to be positive or if the contrary is preferred. If direction is measured, during summation some of the magnitude will be lost as negative and positive differences are added together. Continuing this line of reasoning, t-tests are two-sided ($\alpha = 2 \times 0.025$). There are instances where subscale scores can't be calculated because less than half the responses from the subjects are usable. When the subscale score for half a pair can't be calculated, paired differences also cannot be calculated, changing the number of paired scores used to calculate $t_{observed}$. 

71
A disadvantage from using natural groups is the limited number of subjects. Despite using two sub-organizations, my sample is uncomfortably small. My pairing methodology and attrition caused further reductions. A small sample size reduces the power to detect differences between the two groups and raises the possibility sampling fluctuations could reverse the sample results from population results.

The opportunity to collect real-world data offsets limitations of using natural groups. Subjects in both sub-organizations were generally cooperative. Without a controlled laboratory environment, the effects of the training on perception are real, not hypothetical. Using real subjects, may lend credibility to my results for practitioners considering TQM training in their own organizations.

I like the thought process of Fortune and Hutson (1984) for understanding validity weaknesses in quasi-experiments. Three phenomena that may directly or jointly confuse the validity might be present in my research: unequal outside influences, a priori differences, and an imperfect measurement instrument. It's possible all three conditions exist in my study. Interaction between these three conditions complicates things even more. If all three conditions exist, even the interactions can interact with each other.

There might be unequal effects of outside influences. I tried to compensate for any unevenness by using two sub-organizations at the same level on the hierarchy within a single parent organization, but influences could creep in during the normal course of doing business. The two sub-organizations have similar domains of responsibility in different areas; their work isn't redundant. TQM material could cross the desks of anyone in either group without my knowledge.

There are a priori differences between the groups. By using natural groups, I lost the benefits of random assignment. There is some degree of self-selection of the subjects comprising the groups, particularly for the supervisors, simply because the expertise of each person must contribute to the domain of responsibility for the sub-organization. People possessing common traits are often found in the same discipline or profession. By interest, background, education or some other characteristic, the people in each sub-organization have something in common with others in their group. The two sub-organizations delve into entirely different academic areas. Although the pre-test comparisons indicate no significant differences between the two groups, the two groups cannot be considered statistically equivalent. We know there is self selection, particularly for the supervisors, based on their decision to work in the academic fields of the two sub-organizations.
No measurement instrument is perfect. Responses are based on perceptions, which differ from person to person. It's possible some people may infer they are to predict the responses for the other half of their pair. Using "can" (hypothetical) rather than "does" (actual) wording in the rating options allows more than one interpretation, predicting rather than perceiving.

**Instrument Reliability:** Cronbach's Alpha is a recognized indicator of instrument reliability. A reliability of .70 or higher is desirable (Walker, 1985). I have some confidence in the instrument since each of the calculated values meets this criteria (Leadership: 0.71; Communication: 0.83; Supervision: 0.87). Not all the questions for any of the three subscales could be used in my calculations. A significant portion of the collected data is unusable; there's no way to interpret those responses. The calculation requires the same number of responses for each item used in the calculation. Including the largest number of subjects is desirable, a particularly pressing point with the small number of subjects involved in this study. Only questions receiving no unusable responses from all of the subjects are included in the calculation to incorporate the greatest number of subjects. Unfortunately, not all items for each subscale received no unusable responses: two of eight for Leadership; two of three for Communication; and four of eighteen for Supervision.

**Hypothesis Testing.** I felt my hypothesis testing was most confined by the small sample size. Every person in both sub-organizations fitting my definition of supervisor or subordinate was asked to participate in the study, yet the number of subjects was rather small at the start of the study. When a subject was lost through routine personnel turnover, the participation of two or more subjects could be affected. Since each pair of subjects provides a single datum for analysis, the difference between their subscale scores, the loss of any single subject cost at least two subjects worth of data. I have a hunch more significant differences would have been observed if the sample was larger.

The details of my focus of analysis are contained in my results. Three subscales (Leadership, Communication and Supervision) might be the three most important subscales in terms of the supervisor-subordinate relationship, as my secondary survey indicated, but I want to discuss my hypothesis testing in terms of the design of the study and what I was hoping to find.

**Post-test T-test:** the test I hoped would demonstrate significant differences between the experimental and control groups. I hoped to see that experimental group supervisor-subordinate
paired response differences are significantly smaller than the same response differences for the control group. If the introductory TQM training had no impact on participant's perceptions, the post-test differences for all the subjects should have been indistinguishable. Only the subscale for work unit guidance demonstrated a significant difference. The individual tasks comprising this subscale reflect some of the big-picture, cooperative philosophies of TQM.

**Pre-test T-test:** the test I used to see where the two groups differed prior to treatment. I know the groups weren't statistically equivalent since there was no random assignment, but the more similar they were at the start of the study, the easier the interpretation of the results would be at the conclusion. I was pleased that there were no measured significant differences between experimental and control group supervisor-subordinate paired response differences.

**Pre-test and Post-test versus Post-test Only:** the test I used to determine whether the instrument had an impact on the subjects. I compared the post-test paired differences of subjects taking the pre-test (experimental and control group subjects) with subjects taking only the post-test (experimental and control group subjects). If the instrument had an effect on the subjects, there should be a significant difference between the paired response differences for this arrangement of subjects. Only the subscales for work unit planning and external awareness demonstrated significant differences. Both of these subscales consist of tasks addressed by a TQM approach within an organization. The subscale for work unit planning involves looking ahead and planning. The subscale for external awareness involves an information gathering and sharing approach.

**Pre-test Post-test Differences T-test:** the test I used to determine whether supervisor-subordinate paired response differences are affected by the treatment. This is the most confusing test I performed; it's almost like a dependent t-test except that the differences are computed for each group before the comparison is made. I'm comparing the differences in differences between the experimental and control groups. I use absolute values of all the differences to compare the magnitudes. The interpretation of the results isn't obvious since there are four possible changes affecting any significant difference: experimental group supervisor, experimental group subordinate, control group supervisor, and control group subordinate. A greater difference might be the result of uneven improved evaluations, not the result of less similar perceptions. A lessor difference might be the result of more favorable self evaluations combined with more unfavorable evaluations from subordinates. I think both these changes occurred in the subscales for work unit guidance, program evaluation, leadership, broad perspectives and strategic view. All these
subscles incorporate issues crucial to TQM. On the positive side, when attention is paid to the
tasks comprising these subscles, the result can be beneficial to the organization. On the negative
side, when these tasks are ignored or poorly conducted, the organization suffers. In the subscle
for work unit guidance, cooperation is key. In the subscle for program evaluation, positive
change is the driver. Leadership is the subscle most important to the supervisor-subordinate
relationship. The subscle for broad perspectives implies an understanding of the big-picture. The
subscle for strategic view shows understanding of how the organization fits into the big-picture.

First Time Experimental Group T-test: the test I used to determine whether the effects of the
treatment could be seen within the experimental group. Supervisor-subordinate pairs were
randomly assigned to participate in the pre-test. As much as possible, pre-test participants were
statistically equivalent to post-test only participants. Since work in the experimental sub-
organization continued naturally except for the treatment, I assume any differences can be
attributed to the introductory TQM training. By comparing the pre-test response differences with
the post-test response differences of the post-test only pairs, I hoped to find some indication the
treatment had an impact. Only the subscle for work unit planning demonstrated a significant
difference; this subscle involves looking ahead and planning, two themes of TQM.

Pairs Completing Both the Pre- and Post-tests. An examination of these supervisor-subordinate
pairs provided an opportunity to look at the data a little differently. Dependent t-tests of the
control group responses show no significant differences between pre-test and post-test scores.
Dependent t-tests of experimental group responses showed two subscales (Work Unit Monitoring
and Program Evaluation) having significant differences between the pre-test and post-test.

I used these supervisor-subordinate pairs to try to find a pattern for the unusable responses.
During my analysis I couldn't decipher the reasoning behind the responses. The ratings don't do
this task and can't judge were particularly troublesome because they can't be used for comparison.
Both appeared more often than I anticipated. Supervisors weren't shy about saying they couldn't
judge some aspect of their task performance. Three of five experimental group and two of four
control group supervisors gave themselves this rating at least once. I couldn't find a pattern,
although the subscales for budgeting and material resource administration caused some confusion.
Supervisors who said don't do this task received all possible selections except cannot do this task at
an acceptable level from their subordinates. Each supervisor received an exceptional self rating at
least once in the post-test, which is generally more favorable than their pre-test self rating and
somewhat more favorable than the ratings from their paired subordinates. These observations, while not testing any hypotheses, guided my analysis. Some of the inconsistencies identified weaknesses in my study and others raised issues which are worthy of further research. If I had found consistency, I probably wouldn't have attempted all my analysis approaches.

Supervisors Don't Always Rate Themselves More Favorably Than Their Subordinates Rate Them. This finding was a real surprise. All of the literature I reviewed caused me to expect supervisors to rate themselves more favorably. Supervisors rating themselves more modestly than their subordinates rated them can be seen in all the categories (pre-test, pre-test & post-test and post-test only) for both groups. This finding prompted some modifications to my analysis. Since I could no longer assume subordinate scores subtracted from supervisor scores would be a positive number, I couldn't use one-sided t-tests. I had to use two-sided t-tests of the magnitudes of the differences, changing my critical t-values.

Pairs with the supervisor score less than the subordinate score are most prevalent in the experimental group. I explored this observation using the Wilcoxon-Mann-Whitney test on each subscale. For the pre-test responses, two subscales (Representation and Coordination) differ significantly between the experimental and control groups. For the post-test responses, six subscales (External Awareness, Interpretation, Coordination, Work Unit Monitoring, Leadership and Flexibility) differ significantly between the two groups. Experimental group subordinates tended to give the more favorable scores within their pairs. Control group supervisors tended to give the more favorable scores within their pairs. This is what I hoped to see. If experimental group subordinates understand the complexities of their supervisor's position, they might have an appreciation of their supervisor's task performance. If control group supervisors haven't learned to work cooperatively or share knowledge with others, subordinates might resent their supervisor's actions. While I must be cautious in my interpretation of these findings, it appears the experimental group weathered the university-wide budget cuts better than the control group.

Improvements I would Make

If I were to do this study again, there are a few details I would change. I would have a pilot study. I would use a different evaluation scale for the 118 task items. And I would change the procedures for interacting with the subjects.
Pilot Study. The study would have benefited from a pilot study. There must be differences between federal government and state university employees. Identification of differences in wording interpretation may have averted potential problems (e.g., the budgeting subscale caused a great deal of confusion with the subjects). Both organizations handle paperwork but have different customers. A government organization channels and coordinates primarily within itself and with other government organizations; a university does the same, but has individual customers (students and faculty) too.

None of the research directly preceding my measurement instrument involved the evaluation of training. OPM devised the Management Excellence Inventory to help identify areas supervisors would benefit from training. MSPB used their questionnaire to report on the status of federal supervisor quality. My measurement instrument is directed to the evaluation of introductory TQM training.

TQM is happening in offices across the country but there are virtually no scientific studies of this training. The recommended format for the video library is to view the tapes and then have an open discussion. Perhaps another approach would have been appropriate for this population. My subjects are immersed in a learning environment and many are professional teachers in addition to having administrative responsibilities.

There are many ways to collect data. Research is a routine activity at a university. While the general population might not understand where the data are going or how they will be analyzed, people at a university are exposed to these techniques daily. The procedures for organizing a study on a university campus might differ from other settings if the data collected is to be unbiased.

Although universities have a reputation for openness and freedom of expression, the number of permanent employees working on the campus is fairly small in relation to the number of people on the campus. It's possible the lack of privacy or anonymity might make people more reluctant to be forthright than people in large government or business organizations.

Evaluation Scale. I was astonished to find such disparity between responses. There are instances where a single subject's response to a single question in the pre-test and the post-test seem to contradict. And it's not unusual for paired supervisors and subordinates to say at the same time
that the task isn't performed and that it's performed exceptionally well. After mulling over the apparent inconsistencies, I think the source of the confusion is the evaluation scale.

At first glance, my evaluation scale appears to be a Likert scale, but it isn't. The first and last choices might be indistinguishable to some respondents while they might have entirely separate meanings to others. The interpretation could be subtle enough to change with each reading. The interpretation of the scale shouldn't be fluid and should have a Likert-like structure.

I easily envision respondents pondering whether a task is doable instead of how it's performed, predicting rather than perceiving. The nebulous wording of the rating selections promotes this confusion. Although the question asked for each task is "What is your performance level on this task?", the choices say "cannot" or "can" (theoretical). This uncertain word is confusing. It's possible a person can do a task that isn't part of their current job. The options should have been "doesn't" or "does" (actual). I would also consider changing the measurement term from "acceptable level" to something less subjective.

Procedures. That subjects in my study tended to contradict previous studies of self versus subordinate evaluation is intriguing. From this study I can only speculate the reasons. Infante & Gorden (1971) observed the same relationships between supervisor and subordinate responses when supervisors tried to predict what their subordinates would say. I suspect supervisors and subordinates may have been predicting what the other half of their pair would say. The fact that the control group post-test subjects didn't demonstrate this contradiction with previous evaluation studies to the same degree, strengthens my support for this supposition. The post-test survey distribution had little fanfare. The linking of supervisors and subordinates or the use of the data wasn't emphasized. In the experimental group, participation in the training would have generated interest in the study for the duration of the treatment. In the control group, there was no treatment, only business as usual.

The experiment shouldn't have been so fully explained to the subjects before the study began. There are ethical implications in hiding details from subjects, but a desire to minimize any fear might have changed the experiment. The subjects were briefed before the pre-tests were distributed, the instrument cover page reminded the subjects of the scheme of the study, an elaborate confidential code protected identities, and all the supervisor and subordinate paperwork was paired at distribution times.
If I were to do this study again, I would try to de-emphasize the link between individual supervisors and subordinates. I think the reminders of how the data would be analyzed caused people to predict rather than evaluate. Supervisors might have predicted how their subordinates would evaluate them, perhaps so they wouldn't look foolish to their peers. Subordinates might have predicted how supervisors would evaluate themselves since the supervisors knew their responses would be compared. Both might have feared the other would learn about their response.

I did see some indicators of fear on the part of the subordinates, particularly in the control group. There was a great deal of interest in the confidentiality codes. Some respondents even signed their name across the flap of the envelope used to store their code number for future use so they could detect any signs of tampering.

Recommendations for Managers

I have some recommendations for managers considering TQM training and training evaluation. Unfortunately, based on this study, I cannot say much about the training itself; most of my potential comments are inappropriate since they would be based on other personal experiences. I didn't participate in any of the training sessions nor did I select the video tapes. Identification of training goals is a common topic in the training literature. While I don't think all training must have a concrete objective, the content of any training should be consistent with an organization's culture, or the results of the training might adversely affect the organization's operations. The experimental sub-organization had definite goals for its training venture. The literature urges support from top management. This factor was cited by the experimental group training facilitator as a reason for her belief the experimental sub-organization's TQM training effort was a success. The literature also recommends participation by all members of an organization. The experimental sub-organization followed this advice. Participation was mandatory. The presentation format and content should match the audience. The experimental sub-organization made an informed decision choosing to use Dr. Deming's philosophy and selecting each video tape based on their needs.

If the effect of the training must really be measured, the Solomon Four Group Design works nicely. Even if no control group can be found, a pre-test provides a baseline for measurement. A conscious decision of whether regurgitation of facts and figures or performance on the job reflect acquired knowledge should be made before putting together the measurement instrument. Even in confidential surveys, potential for fear should be addressed. Subjects shouldn't have to fear
retaliation or reprisals for providing honest responses. The logistics of distributing and collecting the measurement instruments should be well-planned. And the measuring process shouldn't be complicated or time-consuming.

Future Research
I think it's fascinating to try to understand what is happening in knowledge worker environments. Since the workforce is becoming more educated and less labor-intensive at the same time forces are squeezing the workers, it will be interesting and painful to see how equilibrium is reached in the workplace.

For any future studies in this area, I urge the researcher to use larger samples. Continuing the comparison of supervisor and subordinate responses is sure to uncover interesting findings. Pursuing any of the subscales in greater depth could provide useful information for practitioners. Longitudinal studies with methodologies similar to mine can either have more data collection points over the course of the training (perhaps necessitating a longer training period) to see the progress of changes or have a data collection point long after the training is completed to determine any long-term effects of the training. Any linking of training with organizational performance measures will be welcomed by practitioners trying to justify training expenditures. And there are many uses for an investigation of supervisors and subordinates predicting evaluations versus assigning evaluation ratings: communication studies, performance studies or satisfaction studies to name a few.

Some research questions which would be interesting to pursue include:
1.) How does introductory TQM training for supervisors and their subordinates change the differences between a supervisor's perspective of that supervisor's task performance and the subordinate's perspective of that supervisor's task performance as measured before training begins, at the end of training and six months after training?
2.) How does introductory TQM training for supervisors and their subordinates change the performance of that organization as measured before training begins and six months after training?
3.) How does the evaluation of supervisors by their subordinates change that organization?
4.) How do the logistical details of gathering paired supervisor and subordinate evaluations of the supervisor affect the evaluations given by the supervisor and subordinate?
5.) How does the evaluation of supervisors by their subordinates affect communication within the dyad?
Chapter 7
CONCLUSION

Chapter 7 is a succinct conclusion to my study of whether introductory TQM training affects the measured differences between supervisor and subordinate perceptions of supervisor task performance.

I cannot say that introductory TQM training causes supervisors and subordinates to become more similar in their perceptions of supervisor task performance. The results of my hypothesis testing are inconclusive. I was disappointed, although not surprised, to find few significant differences between the experimental and control groups after the training program was completed. Small samples have limited power in discerning differences. In each hypothesis test where I hoped to see differences, I was heartened to find at least one subscale with significant differences between experimental and control group paired supervisor-subordinate response differences. When I examined the two groups prior to training, no significant differences were measured. Even if the differences of every subscale had supported my hypotheses, the interpretation would have been unclear since small samples may actually be reversed from the true population.

All the subscales exhibiting a significant difference at least once during my hypothesis testing are comprised of tasks affected when a TQM philosophy is guiding an organization.

I'm glad I had the opportunity to use Solomon's Four Group Design. A little care when a study is in the planning stages pays off with the versatility of the analysis. Comparison testing of the data is easily envisioned, permitting different comparisons of the data for a multifaceted evaluation of a training effort. Skeptics can perform dependent analysis to gain greater understanding of the responses.

Investigation of directional comparisons of the paired responses seems to indicate the experimental group weathered the university-wide budget cuts less negatively than the control group. Experimental group subordinates had a more favorable view of their supervisors than both their control group counterparts and their paired supervisors at the conclusion of the treatment.
REFERENCES


APPENDIX A

THE INSTRUMENT*

This survey asks you to share your experiences and opinions concerning your/your supervisor's/the faculty member's role. In order to obtain a comprehensive assessment, similar questionnaire packages are being distributed to both supervisors/faculty members and support staff members. Although the information will be confidential, the links for each pair must be maintained so that a thorough analysis can be performed. None of the information will be used to identify any individuals, but it is critical the opscan sheets be properly numbered and individuals complete the appropriate questionnaire package.

For this study, only a limited number of identified relationships are being examined. Although other relationships do exist, only those included in this study.

You should answer the questions with respect to your/your supervisor's/the faculty member's job and performance. For questions 1 through 118, the format will be:

What is your/your supervisor's/the faculty member's performance level on this task?
1 = doesn't do this task
2 = cannot do this task at an acceptable level
3 = can do this task at a barely acceptable level
4 = can do this task at an acceptable level
5 = can do this task at an above average level
6 = can do this task exceptionally well
7 = I can't judge

For the remainder of the questions, the format will be less structured.

Instructions regarding the numbering scheme and defining the subjects for each version of the survey.

Instructions regarding the collection of the completed questionnaires and opscan sheets.

The importance of correctly linking the relationships between the supervisor/faculty member and the support staff cannot be overemphasized. The survey content is noticeably lacking demographic details to lend credibility to the assurance that no attempt will be made to identify any individuals.

Please complete the survey without consulting any coworkers or outsiders

*Modified to: reflect the variations for each subject category; eliminate unnecessary details; and address only the questions analyzed for this thesis.
WHAT IS YOUR/SUPERVISOR'S/FACULTY MEMBER'S PERFORMANCE LEVEL ON THIS TASK?
1 = DOESN'T DO THIS TASK
2 = CANNOT DO THIS TASK AT AN ACCEPTABLE LEVEL
3 = CAN DO THIS TASK AT A BARELY ACCEPTABLE LEVEL
4 = CAN DO THIS TASK AT AN ACCEPTABLE LEVEL
5 = CAN DO THIS TASK AT AN ABOVE AVERAGE LEVEL
6 = CAN DO THIS TASK EXCEPTIONALLY WELL
7 = I CAN'T JUDGE

Work Unit Planning
1. Develops plans for work unit that includes costs, personnel or material needs (for example, supplies)
2. Estimates the resources (staff time, costs) required to complete a job
3. Develops work unit plans that extend beyond one year
4. Determines whether proposed actions are technically workable

Work Unit Guidance
5. Establishes priorities among work unit activities and projects
6. Schedules work so that available resources are used most effectively
7. Provides specific guidance on how achieving work objectives will be measured
8. Prepares technical procedures used in work unit
9. Develops and implements procedures to keep work unit running smoothly
10. Considers workload demands in approving leave and overtime

Budgeting
11. Prepares or provides input into work unit's budget
12. Uses financial reports in managing work unit
13. Estimates financial needs beyond current budget year
14. Keeps track of work unit expenses and monitors against budget

Material Resources Administration
15. Oversees or participates in managing work done by contractors
16. Applies contract and procurement rules and regulations in managing work unit
17. Sees that enough supplies and equipment are available to do the job
18. Oversees equipment maintenance
19. Prepares justifications for equipment or other materials needed by work unit

Supervision
20. Helps employees identify their developmental needs and gets appropriate training or experience
21. Gives praise for good work
22. Recognizes when employees are having difficulty performing work
23. Maintains a balance between a concern for people and concern for productivity
24. Explains tasks so that employees clearly understand their duties
25. Coaches and counsels employees on conduct (about behavior) not related to performance
26. Coaches and counsels employees on a timely basis about their performance
27. Monitors time and attendance and sick leave usage to prevent abuse
28. Provides or arranges for the orientation of new employees
29. Gives on-the-job training to employees (or direct others to do so)
30. Keeps adequate employee records and reports (e.g., accident, injury)
31. Makes daily work assignments to employees
32. Monitors and documents employee performance
33. Assigns work to employees based on individual capabilities
34. Allows employees to work without unnecessarily close supervision
35. Makes sure that employees have safe working conditions
36. Assists employees in solving technical problems they have with assigned tasks
37. Does journeyman work in an emergency
WHAT IS YOUR/SUPERVISOR'S/FACULTY MEMBER'S PERFORMANCE LEVEL ON THIS TASK?
1 = DOESN'T DO THIS TASK
2 = CANNOT DO THIS TASK AT AN ACCEPTABLE LEVEL
3 = CAN DO THIS TASK AT A BARELY ACCEPTABLE LEVEL
4 = CAN DO THIS TASK AT AN ACCEPTABLE LEVEL
5 = CAN DO THIS TASK AT AN ABOVE AVERAGE LEVEL
6 = CAN DO THIS TASK EXCEPTIONALLY WELL
7 = I CAN'T JUDGE

Personnel Management
38. Develops appropriate performance elements and standards
39. Uses performance elements and standards to assess employee performance and give feedback
40. Uses personnel management practices that support agency EEO objectives
41. Applies personnel rules and regulations to selection and promotion actions and decisions
42. Uses personnel management practices that promote good labor-management relations (e.g., resolving grievances at an informal level)
43. Initiates recommendations for awards and discipline (written)
44. Enforces employee compliance with required procedures (e.g., time keeping, safety procedures)

External Awareness
45. Keeps up-to-date with latest technology, methods, and equipment relevant to the work unit
46. Keeps up-to-date with the overall structure and functions of related organizations within the agency
47. Keeps up-to-date with regulations, policies and administrative procedures affecting work unit
48. Maintains currency in technical knowledge required by job

Interpretation
49. Identifies and applies information from higher management that affects the work unit
50. Explains regulations, policies, rules and procedures to employees
51. Prepares letters, memos or reports that reflect higher management policy and directives
52. Keeps employees informed of changes in procedures, policies and rules that affect their work

Representation
53. Responds to inquiries and requests from outside the work unit
54. Explains work unit projects or activities to nonexperts in terms they can understand
55. Discusses work unit issues (e.g., work status, resource needs, employee concerns) with higher level management or other persons in the organization

Coordination
56. Maintains good working relationships with immediate supervisor and peer supervisors
57. Keeps supervisor informed of problems and work status
58. Coordinates with other units to promote smoother operations and to maintain schedules

Work Unit Monitoring
59. Adjusts to changes in workload, resources or priorities
60. Spots irregularities in work unit operations before major problems occur
61. Observes work in process to ensure that jobs are completed on time and are of high quality
62. Documents workflow and results
63. Reviews work in different parts of work unit to ensure coordination

Program Evaluation
64. Identifies ways to improve the efficiency and effectiveness of work unit operations
65. Critically and realistically assesses the overall effectiveness of the work unit
66. Uses project or job reviews or other information to improve work unit operations
67. Assesses progress toward achieving work unit objectives
68. Identifies ways to improve work unit procedures and methods

Communication
69. Speaks clearly and effectively
70. Presents ideas and facts clearly and effectively in writing
71. Listens to others and shows understanding of what they are saying.
WHAT IS YOUR/SUPERVISOR'S/FACULTY MEMBER'S PERFORMANCE LEVEL ON THIS TASK?
1 = DOESN'T DO THIS TASK
2 = CANNOT DO THIS TASK AT AN ACCEPTABLE LEVEL
3 = CAN DO THIS TASK AT A BARELY ACCEPTABLE LEVEL
4 = CAN DO THIS TASK AT AN ACCEPTABLE LEVEL
5 = CAN DO THIS TASK AT AN ABOVE AVERAGE LEVEL
6 = CAN DO THIS TASK EXCEPTIONALLY WELL
7 = I CAN'T JUDGE

Interpersonal Sensitivity
72. Is consistent and fair in dealing with employees
73. Provides negative feedback in a constructive manner
74. Accurately assesses the strengths and weaknesses of others
75. Realistically assesses own/supervisor's/faculty member's) strengths and weaknesses and impact on others
76. Accepts and makes use of justified criticism
77. Achieves objectives by discreetly using power or authority
78. Resolves differences through informal discussions or counseling
79. Handles problems diplomatically
80. Is accessible to employees

Leadership
81. Actively promotes cooperation and teamwork within work unit
82. Supports higher management and policy decisions with employees
83. Demonstrates a positive approach to employees - a "can do" attitude
84. Makes "tough" or unpopular decisions
85. Accepts responsibility for work unit and doesn't "pass the buck"
86. Shows respect and support for employees
87. Sets a good example for employees
88. Conducts meetings in a way that achieves desired results

Flexibility
89. Handles job pressures and stress
90. Encourages open communication and input from employees
91. Admits to and learns from mistakes
92. Revises priorities and procedures when new information suggests a change is needed
93. Is flexible in dealing with different situations and people
94. Handles more than one problem at a time
95. Encourages employees to be innovative and creative in dealing with work situations
96. Tries new ideas and methods to get the job done as effectively and efficiently as possible

Action Orientation
97. Manages own (supervisor's) time efficiently
98. Takes action rather than waiting to react to actions as they occur
99. Acts decisively on own authority when timely action is needed
100. Develops and implements solutions to problems that affect work unit
101. Takes the initiative in gathering information needed to get the work done

Results Focus
102. Works persistently toward a goal despite opposition, distractions and setbacks
103. Is concerned with achieving final results as well as conducting day-to-day activities
104. Does what is necessary to get the job done
105. Sets challenging but realistic deadlines for completing work

Broad Perspectives
106. Maintains a sense of mission in day-to-day activities
107. Takes into account a wide variety of factors that affect work unit efficiency and effectiveness
108. Takes longer-term goals into account while preparing short-term plans and schedules
109. Defines the "big picture" to employees (how their jobs relate to others, etc.)
WHAT IS YOUR/YOUR SUPERVISOR'S/FACULTY MEMBER'S PERFORMANCE LEVEL ON THIS TASK?
1 = DOESN'T DO THIS TASK  
2 = CANNOT DO THIS TASK AT AN ACCEPTABLE LEVEL  
3 = CAN DO THIS TASK AT A BARELY ACCEPTABLE LEVEL  
4 = CAN DO THIS TASK AT AN ACCEPTABLE LEVEL  
5 = CAN DO THIS TASK AT AN ABOVE AVERAGE LEVEL  
6 = CAN DO THIS TASK EXCEPTIONALLY WELL  
7 = I CAN'T JUDGE

Strategic View
110. Recognizes discrepancies and deficiencies in various types of information
111. Is logical and systematic in analyzing problems and issues
112. Identifies patterns in events or information
113. Recognizes the key parts of a problem or issue
114. Gathers and shares information through informal inquiry and discussion
115. Recognizes when to take action and when to "bide time"

Environmental Sensitivity
116. Demonstrates awareness of sensitive organizational policies and activities
117. Considers the ethical implications of a given course of action
118. Understands the importance of non-technical factors (e.g., funding, special interests) in higher management decisions

How would you rate yourself/your supervisor/the faculty member on the following abilities?
119. Ability to make efficient use of time
   1 = Very low work output; perform generally at an unsatisfactory pace
   2 = Low work output; perform at a slower than acceptable pace
   3 = Good work output; perform at an acceptable pace
   4 = High work output; perform at a fast pace
   5 = Very high work output; perform at an unusually fast pace
   6 = Don't know/Can't judge

120. Ability to do quality work which meets acceptable standards
   1 = Quality of work is usually unacceptable and does not meet minimum standards
   2 = Quality of work is usually somewhat inferior
   3 = Quality of work is acceptable but usually not superior
   4 = Quality of work is usually superior
   5 = Quality of work is always the highest
   6 = Don't know/Can't judge

121. Ability to handle multiple job operations
   1 = Cannot efficiently perform a limited number of different assignments
   2 = Can efficiently perform a limited number of different assignments
   3 = Can efficiently perform an acceptable number of different assignments
   4 = Can efficiently perform many different assignments
   5 = Can efficiently perform an unusually large variety of different assignments
   6 = Don't know/Can't judge

122. Ability to propose, produce or apply new or better approaches to products, processes or services
   1 = Work shows little or no creativity; make little or no use of new or better approaches
   2 = Work shows limited creativity; make only limited use of new and better approaches
   3 = Work shows a moderate amount of creativity; make moderate use of new and better approaches
   4 = Work shows much creativity; make much use of new and better approaches
   5 = Work shows exceptional creativity; make exceptional use of new and better approaches
   6 = Don't know/Can't judge
123. Ability to perform a wide range of tasks appropriate for the job
   1 = Have few knowledges or skills required for the job; able to perform only the simplest tasks
   2 = Have some knowledges or skills relevant to the job; able to perform typical tasks, but only
      a few more demanding tasks
   3 = Have many knowledges or skills relevant to the job; able to perform most tasks, but only
      some of the most demanding
   4 = Have almost all necessary and desirable knowledges and skills; able to perform almost all
      tasks
   5 = Have all necessary and desirable knowledges and skills; able to perform all tasks
   6 = Don't know/Can't judge

124. Frequency of providing feedback to staff
   1 = Yearly; when performance evaluations are completed
   2 = When a job is either particularly well done or there is a big problem
   3 = Every week; a regular meeting is scheduled
   4 = Daily; a regular meeting is scheduled
   5 = Constantly; all work is directed and controlled
   6 = Don't know/Can't judge

125. When you meet with your support staff/your supervisor/the faculty member, how effective are these
     meetings?
   1 = Unproductive; cause confusion or dissent
   2 = Inefficient; too much time wasted
   3 = Somewhat helpful for big jobs
   4 = Very useful
   5 = Indispensable; couldn't be as productive without them
   6 = Don't know/Can't judge

126. Considering all relevant factors, to what extent do(es) you/your supervisor/the faculty member
     contribute to the accomplishment of the organization's objectives goals and missions?
   1 = Overall contribution is rarely positive
   2 = Overall contribution is positive but not substantial
   3 = Overall contribution is positive and usually meets expectations
   4 = Overall contribution frequently exceeds expectations
   5 = Overall contribution typically exceeds expectations and serves as a leading example for
      others
   6 = Don't know/Can't judge

127. Considering all relevant factors, to what extent do you contribute to the accomplishment of the
     organization's objectives goals and missions?
   1 = Overall contribution is rarely positive
   2 = Overall contribution is positive but not substantial
   3 = Overall contribution is positive and usually meets expectations
   4 = Overall contribution frequently exceeds expectations
   5 = Overall contribution typically exceeds expectations and serves as a leading example for
      others
   6 = Don't know/Can't judge

128. How would you rate your/your supervisor's/the faculty member's overall quality/effectiveness?

<table>
<thead>
<tr>
<th>NOT AT ALL EFFECTIVE</th>
<th>NEITHER EFFECTIVE NOR INEFFECTIVE</th>
<th>VERY EFFECTIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

129. How many separate opscan forms have you completed as part of this survey?
   1 = This is the first time I have seen this survey
   2 = I participated in the initial survey in September - this is the second opscan form I've done
   3 = I participated in the initial survey in September - this is the third opscan form I've done
   4 = I participated in the initial survey in September - this is the fourth opscan form I've done
Appendix B
FOURTEEN TASKS IMPORTANT TO THE SUCCESS OF A SUPERVISOR
From
Federal First-Line Supervisors: How Good are They
A 1992 U.S. Merit Systems Protection Board Report

1.) Establish priorities among work unit activities and projects.

2.) Schedule work so that available resources are used most efficiently.

3.) Use performance elements and standards to assess employee performance and give feedback.

4.) Explain tasks so that employees clearly understand their duties.

5.) Maintain good working relationships with immediate supervisor and peer supervisors.

6.) Keep supervisor informed of problems and work status.

7.) Speak clearly and effectively.

8.) Listen to others and show understanding of what they are saying.

9.) Be consistent and fair in dealing with employees.

10.) Actively promote cooperation and teamwork within work unit.

11.) Demonstrate a positive approach to employees - a "can do" attitude.

12.) Show respect and support for employees.

13.) Set a good example for employees.

14.) Manage own time efficiently.
Appendix C
INTERVENTION SCHEDULE

Week | Training |
-----|----------|
1    | Video - *The New Economic Age*
     | Discussion - How do these ideas relate to us? |
2    | Video - *The Quality Leader*
     | Discussion - Which beneficial ideas can we adopt for our office? |
3    | Video - *Quality in the Office*
     | Discussion - What does quality in our office mean? |
4    | Workshop - Flowcharting |
5    | Video - *How Managers and Workers Can Change*
     | Discussion - What ideas are pertinent to our situation? |
     | Concerns about flowcharting progress. |
6    | Video - *What to Do Instead of Performance Appraisals*
     | Discussion - Problems in identifying process steps or flows for flowcharting. |
7    | Video - *Understanding Profound Knowledge*
     | Discussion - Where is the sub-organization in having profound knowledge about its systems and world? |
8    | Video - *The Funnel Experiment*
     | Discussion - Determine problem-solving needs and how we can avoid tampering. |
9    | Video - *The Fourteen Points*
     | Discussion - Which points do we need to address first? |
10   | Video - *Cooperation, The Key to Quality*
     | Discussion - Report progress of problem-solving teams. |
11   | Video - *Competition, Cooperation and the Individual*
     | Discussion - Identify areas for improvement in our office.
Appendix D

EXPERIMENTAL GROUP TRAINING FACILITATOR INTERVIEW SUMMARY

This is not a transcript.

Are you responsible for the training and development of the staff?
No

How did you decide to put together a training program for the sub-organization?
The top manager, others in the group, and she attended a 4 day Deming seminar a few months before the training began. She was asked to look into training for the whole sub-organization.

Why did you choose to use Deming's philosophy?
They liked what they saw at the workshop. Based on his definitions, terms, and concepts, he seemed to be more philosophical and more academic than some of the other total quality gurus.

How did you decide to purchase tapes from the Deming library?
She heard about the company that sells the tapes. It was too expensive to send everybody to a Deming workshop. She and the top manager thought the videos were the next best thing.

How did you decide which tapes to use?
She chose the tapes based on what she knows about the staff and what she knows about Deming.

Were your selections limited by a budget?
Somewhat, but a considerable sum was spent purchasing the tapes.

So what was the limit based on...?
Time. We didn't have time to watch 25 films. Also, some of the tapes didn't pertain to the sub-organization as they have to do with industry.

How did you decide you had enough training material? Was time the total constraint? Did you have a feeling you covered everything you wanted to cover?
The sub-organization was playing it by ear, starting with 10 or 12 tapes. Their selections turned out to be OK. They felt the tapes chosen were enough for background.

What were the goals of the training?
To help people understand the need to look at work from a systems point of view. The training was to help people appreciate how their work affected other people's work, to encourage the office to be more customer oriented, and to help get through some budget cuts in a smart way (thinking through how to restructure the office and responsibilities and improve work at the same time).

Were you hoping to see a change in activity, communication or what?
The sub-organization was hoping to see more collaboration in working together to provide better service for customers during the time budgets were cut and expenses were going to have to be cut. Management hoped to gain better service without crushing morale during the budget cut.

Not only were you looking for more collaboration between the staff and the support people, but you were hoping for more interaction among the peers, among each other?
Yes.
How did you decide on the format?
They followed the format suggested in the booklet that come with the tapes: watch each tape before having an open discussion. The booklet even suggests what to discuss.

Why was she the facilitator rather than another person who attended the seminar?
She knew more about Deming, knew more about total quality, and had experience as a facilitator. She's perceived as a champion of quality improvement in the office.

Was it difficult persuading people to participate in the training?
The meetings were mandatory, you had to go.

Did people participate or were they hesitant?
She made sure everybody participated, but some were more vocal than others. Everybody had a chance to speak.

How did the discussions compare with your expectations?
The discussions were better than she expected. There were some really good discussions about where the office was headed. After the videos were completed, people continued to meet to teach each other about responsibilities. A lot of misconceptions were eliminated and there was greater appreciation for each other's work. She saw reduced tension, increased camaraderie and more cooperation. She spent a lot of time drawing out the people who were inhibited. Whether people were telling the truth during the discussions, who knows?

Could you see a difference between people who participated in the pretest and those who didn't?
No.

Anything else to add?
One interesting thing: they were the only office on campus at the time doing the focused training. Another office was doing a less rigorous training effort. She wonders if there was a bit of Hawthorne effect? People were pretty excited about the experiment because it's highly unusual to have full participation of faculty and support staff. She thinks the training was a success was because of the top manager's commitment, interest and excitement.

So you are glad you did it?
Yes. She doesn't think some people ever bought into it. When the top manager left the sub-organization, they said it never worked. But she doesn't think those people ever understood the concepts of continuous improvement. Not everybody was converted.
Appendix E
SECONDARY SURVEY INSTRUMENT

Please rank the following topics in order of importance as related to a supervisor and subordinate relationship. The topic you consider most important should be rated '10'; the next most important topic should be rated '9', etc. Consider only the ten topics you consider most important:

a. Guidance (schedules, priorities, procedures, activities)
b. Supervision (feedback, explanations, training, assignments, working conditions, sick leave)
c. Personnel Management (PPE, promotions, performance standards, labor relations, evaluations)
d. External Awareness (up-to-date with technology, organization structure and policies)
e. Interpretation (explains policies & procedures)
f. Coordination
g. Work Unit Monitoring (adjusts to changes, spots irregularities, timelines, output quality)
h. Communication (spoken & written clarity, listens)
i. Interpersonal Sensitivity (fair, consistent, diplomatic, constructive with negative feedback, accessible)
j. Leadership (promotes teamwork, accepts responsibility, shows respect, sets good example, "can do" attitude)
k. Action Orientation (proactive rather than reactive, gathers needed information, acts decisively)
l. Results Focus (persists, challenging but realistic goals)
m. Broad Perspectives (long-term goals reflected in short-term plans, explains the "big picture")
n. Strategic View (recognizes information discrepancies, logical, systematic, shares information, recognizes key parts of issues, knows how to "close time")
### Appendix F

#### SECONDARY SURVEY RESULTS

<table>
<thead>
<tr>
<th>SUBSCALE Number and Name</th>
<th>INSTRUMENT Questions</th>
<th>MINOR SURVEY RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># blanks</td>
<td>avg. rank</td>
</tr>
<tr>
<td>Guidance</td>
<td>5-10</td>
<td>2</td>
</tr>
<tr>
<td>Supervision*</td>
<td>20-37</td>
<td>1</td>
</tr>
<tr>
<td>Personnel Management</td>
<td>38-44</td>
<td>8</td>
</tr>
<tr>
<td>External Awareness</td>
<td>45-48</td>
<td>7</td>
</tr>
<tr>
<td>Interpretation</td>
<td>49-52</td>
<td>11</td>
</tr>
<tr>
<td>Coordination</td>
<td>56-58</td>
<td>12</td>
</tr>
<tr>
<td>Work Unit Monitoring</td>
<td>59-63</td>
<td>7</td>
</tr>
<tr>
<td>Communication*</td>
<td>69-71</td>
<td>1</td>
</tr>
<tr>
<td>Interpersonal Sensitivity</td>
<td>72-80</td>
<td>2</td>
</tr>
<tr>
<td>Leadership*</td>
<td>81-88</td>
<td>0</td>
</tr>
<tr>
<td>Action Orientation</td>
<td>97-101</td>
<td>9</td>
</tr>
<tr>
<td>Results Focus</td>
<td>102-105</td>
<td>10</td>
</tr>
<tr>
<td>Broad Perspectives</td>
<td>106-109</td>
<td>4</td>
</tr>
<tr>
<td>Strategic View</td>
<td>110-115</td>
<td>10</td>
</tr>
</tbody>
</table>

* denotes subscale used to focus my analysis.
1 Subscale numbering is consistent throughout my thesis.
2 Subscale name as it appears on the two survey instruments.
3 Of the 21 respondents, this number didn't rank this subscale.
4 Divided the sum of the assigned ranks by the number of written ranks.
5 Divided the sum of the assigned ranks by 21.
Appendix G
LIST OF INSTRUMENT SUBSCALES

<table>
<thead>
<tr>
<th>SUBSCALE NUMBER</th>
<th>SUBSCALE NAME</th>
<th>QUESTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Work Unit Planning</td>
<td>1-4</td>
</tr>
<tr>
<td>2</td>
<td>Work Unit Guidance</td>
<td>5-10</td>
</tr>
<tr>
<td>3</td>
<td>Budgeting</td>
<td>11-14</td>
</tr>
<tr>
<td>4</td>
<td>Material Resources Administration</td>
<td>15-19</td>
</tr>
<tr>
<td>5</td>
<td>Supervision</td>
<td>20-37</td>
</tr>
<tr>
<td>6</td>
<td>Personnel Management</td>
<td>38-44</td>
</tr>
<tr>
<td>7</td>
<td>External Awareness</td>
<td>45-48</td>
</tr>
<tr>
<td>8</td>
<td>Interpretation</td>
<td>49-52</td>
</tr>
<tr>
<td>9</td>
<td>Representation</td>
<td>53-55</td>
</tr>
<tr>
<td>10</td>
<td>Coordination</td>
<td>56-58</td>
</tr>
<tr>
<td>11</td>
<td>Work Unit Monitoring</td>
<td>59-63</td>
</tr>
<tr>
<td>12</td>
<td>Program Evaluation</td>
<td>64-68</td>
</tr>
<tr>
<td>13</td>
<td>Communication</td>
<td>69-71</td>
</tr>
<tr>
<td>14</td>
<td>Interpersonal Sensitivity</td>
<td>72-80</td>
</tr>
<tr>
<td>15</td>
<td>Leadership</td>
<td>81-88</td>
</tr>
<tr>
<td>16</td>
<td>Flexibility</td>
<td>89-96</td>
</tr>
<tr>
<td>17</td>
<td>Action Orientation</td>
<td>97-101</td>
</tr>
<tr>
<td>18</td>
<td>Results Focus</td>
<td>102-105</td>
</tr>
<tr>
<td>19</td>
<td>Broad Perspectives</td>
<td>106-109</td>
</tr>
<tr>
<td>20</td>
<td>Strategic View</td>
<td>110-115</td>
</tr>
<tr>
<td>21</td>
<td>Environmental Sensitivity</td>
<td>116-118</td>
</tr>
</tbody>
</table>
Appendix H
THE ONLY INSTRUMENT ITEMS RECEIVING NO UNUSABLE RESPONSES

<table>
<thead>
<tr>
<th>SUBSCALE</th>
<th>TASK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervision</td>
<td>22 Recognizes when employees are having difficulty performing work.</td>
</tr>
<tr>
<td></td>
<td>23 Maintains balance between concern for people and productivity.</td>
</tr>
<tr>
<td></td>
<td>24 Explains tasks clearly so employees clearly understand their duties.</td>
</tr>
<tr>
<td></td>
<td>34 Allows employees to work without unnecessarily close supervision.</td>
</tr>
<tr>
<td>Communication</td>
<td>69 Speaks clearly and effectively.</td>
</tr>
<tr>
<td></td>
<td>71 Listens to others and shows understanding of what they are saying.</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>78 Resolves differences through informal discussions or counseling.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>79 Handles problems diplomatically.</td>
</tr>
<tr>
<td></td>
<td>80 Is accessible to employees.</td>
</tr>
<tr>
<td>Leadership</td>
<td>83 Demonstrates a positive approach to employees - &quot;can do&quot; attitude.</td>
</tr>
<tr>
<td></td>
<td>86 Shows respect and support for employees.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>90 Encourages open communication and input from employees.</td>
</tr>
<tr>
<td></td>
<td>93 Is flexible in dealing with different situations and people.</td>
</tr>
<tr>
<td></td>
<td>94 Handles more than one problem at a time.</td>
</tr>
<tr>
<td></td>
<td>95 Encourages employees to be innovative, creative in work situations.</td>
</tr>
<tr>
<td>Action Orientation</td>
<td>101 Takes initiative in gathering information needed to get work done.</td>
</tr>
<tr>
<td>Results Focus</td>
<td>103 Concerned with final results as well as day-to-day activities.</td>
</tr>
<tr>
<td></td>
<td>104 Does what is necessary to get the job done.</td>
</tr>
<tr>
<td>Strategic View</td>
<td>110 Recognizes discrepancies, deficiencies in types of information.</td>
</tr>
<tr>
<td></td>
<td>111 Logical, systematic in analyzing problems or issues.</td>
</tr>
<tr>
<td></td>
<td>113 Recognizes key parts of problem or issue.</td>
</tr>
<tr>
<td></td>
<td>114 Gathers and shares information through informal inquiry, discussion.</td>
</tr>
</tbody>
</table>
VITA

I was awarded an MBA from Virginia Polytechnic Institute and State University in 1994; a BS. in Mechanical Engineering from Clarkson College of Technology in 1981; and two AS. degrees, one in Math and Computer Science, the other in Engineering Science, from Jamestown Community College in 1982 and 1979 respectively. I worked as an electronics engineer for US Government defense organizations until beginning my graduate studies in the fall of 1990. My suspicion that the delineation between employee development, business and engineering is uncertain has been substantiated by what I have learned since returning to school.