Pepperdine University
Graduate School of Education and Psychology

THE RELATIONSHIP OF MULTI-RATER FEEDBACK TO TRADITIONAL
PERFORMANCE APPRAISALS

A dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Education in Organization Change

by
Tracy M. Maylett

November, 2005
This dissertation, written by

Tracy M. Maylett

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

October 21, 2005

Faculty Committee:

Kent Rhodes, Ed.D., Chairperson

Susan Nero, Ph.D.

Robert Canady, D.B.A.

Chester H. McCall, Ph.D.
Associate Dean

Margaret J. Weber, Ph.D.
Dean
#TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>viii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>VITA</td>
<td>x</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xii</td>
</tr>
<tr>
<td>Chapter One: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Problem Statement</td>
<td>4</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>5</td>
</tr>
<tr>
<td>Research Questions</td>
<td>5</td>
</tr>
<tr>
<td>Significance of this Study</td>
<td>6</td>
</tr>
<tr>
<td>Conceptual Approach</td>
<td>6</td>
</tr>
<tr>
<td>Definitions and Terms</td>
<td>7</td>
</tr>
<tr>
<td>Chapter Two: Review of the Literature</td>
<td>9</td>
</tr>
<tr>
<td>The Need for Individual Employee Development</td>
<td>9</td>
</tr>
<tr>
<td>The Importance of Feedback to Development</td>
<td>11</td>
</tr>
<tr>
<td>Feedback and Performance Management</td>
<td>19</td>
</tr>
<tr>
<td>The Emergence of Multi-rater Assessment</td>
<td>23</td>
</tr>
<tr>
<td>Variance in Rater Perception</td>
<td>26</td>
</tr>
<tr>
<td>Multi-rater Assessments for Development vs. Appraisal</td>
<td>34</td>
</tr>
<tr>
<td>Multi-rater Assessment and Performance</td>
<td>41</td>
</tr>
<tr>
<td>Summary</td>
<td>45</td>
</tr>
<tr>
<td>Chapter Three: Methodology</td>
<td>47</td>
</tr>
<tr>
<td>Research Design</td>
<td>47</td>
</tr>
<tr>
<td>Research Questions</td>
<td>47</td>
</tr>
<tr>
<td>Research Hypotheses</td>
<td>48</td>
</tr>
<tr>
<td>Research Background</td>
<td>48</td>
</tr>
<tr>
<td>Assessment Instruments</td>
<td>50</td>
</tr>
<tr>
<td>Assessment Administration</td>
<td>54</td>
</tr>
<tr>
<td>Data Collection Procedures</td>
<td>55</td>
</tr>
<tr>
<td>Human Subjects Considerations</td>
<td>56</td>
</tr>
<tr>
<td>Statistical Treatment of the Data</td>
<td>57</td>
</tr>
<tr>
<td>Summary</td>
<td>59</td>
</tr>
<tr>
<td>Chapter Four: Results</td>
<td>60</td>
</tr>
<tr>
<td>Research Questions and Hypotheses</td>
<td>60</td>
</tr>
<tr>
<td>Assessment Instruments</td>
<td>61</td>
</tr>
<tr>
<td>Analysis Overview</td>
<td>63</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Mean, Median, Mode, and Standard Deviation ..................................................... 67
Table 2: Ranges of Multi-rater Overall Scores vs. Performance Evaluation Overall Scores ....................................................................................................................... 67
Table 3: Scale for Interpretation of Correlation Coefficients .............................................. 70
Table 4: Correlation Coefficients (r) and Coefficients of Determination between Multi-rater and Performance Appraisal Categories ........................................................ 71
Table 5: Spearman Rank-order Correlation: Multi-rater Assessment and Performance Appraisal .................................................................................................................. 74
Table 6: Top 10% and Bottom 10% of Multi-rater Rankings Compared to the Same Individuals’ Performance Evaluation Rankings ................................................... 75
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean, median, and mode for performance appraisal and multi-rater assessments</td>
</tr>
<tr>
<td>2</td>
<td>Multi-rater feedback scores plotted against performance evaluation scores (sorted by coded employee number)</td>
</tr>
<tr>
<td>3</td>
<td>Multi-rater feedback scores plotted against performance evaluation scores (sorted by highest to lowest multi-rater scores)</td>
</tr>
<tr>
<td>4</td>
<td>Category summary mean scores for CMT's ranking in the top 10% of multi-rater scores</td>
</tr>
<tr>
<td>5</td>
<td>Category summary mean scores for CMT's ranking in the bottom 10% of multi-rater scores</td>
</tr>
</tbody>
</table>
DEDICATION

To my wife, Lindy, and four sons, Andy, Cameron, Taylor, and Nicholas—my eternal friends, and my world.
ACKNOWLEDGEMENTS

As I my journey began, it soon became apparent that I was not sailing solo. Whether by coercion or simply through love, others quickly came aboard. To each of you, I express my love, admiration, and sincere appreciation.

My wife, Lindy, made my dream her dream— or at least she made me think so. She went far beyond simply seeing my vision, and provided continual support, encouragement, and love. My four boys were always close behind. I realize now that they really know nothing different than Dad going to school. It is because of my love for you that the journey began... and continued. To my family... thank you. I love you.

I was fortunate to belong to a group of school colleagues that provided sincere friendship and drive. I also thank the Pepperdine EDOC faculty and administration, who treated me like a friend and a colleague. I appreciate their wisdom and guidance.

In particular, I wish to acknowledge Dr. Kent Rhodes, Dr. Robert Canady, and Dr. Susan Nero, who served as my committee. The selection of a committee was not something I took lightly, and each of these truly good people has impacted my life.

My friends and associates at work, the DecisionWise team, provided the inspiration to match my academic interests with practical business perspective. I particularly owe a great deal to my business partner, Juan Riboldi, without whom this research would not have taken place.

To my parents, who instilled within me a love of learning and the confidence to work to get things done, I owe my love and gratitude.

Finally, I owe eternal gratitude to my Heavenly Father and His Son. I am mindful that where much is given, much is required... and I truly have been given much.

Thank you.
VITA
Tracy M. Maylett

EDUCATION

Pepperdine University  Doctor of Education in Organization Change  2005  
Brigham Young University  Master of Business Administration  1997  
Utah State University  Bachelor of Arts- Secondary Education; Spanish  1991  
Snow College  Associate of Arts  1988  

PROFESSIONAL EXPERIENCE

DecisionWise  President and Managing Partner  2003-present  
Provo, UT  
Modus Media International  Vice President, Organization Effectiveness  2001-2003  
Boston, MA  
Modus Media International  Vice President, Global Standard Practices  2000-2001  
Boston, MA  
Modus Media International/Stream International  Director, Global Organization Development and Chancellor, Modus Media International University  1997-2000  
Lindon, UT  
Lindon, UT  
Brigham Young University  Part-time Faculty- Department of Organization Leadership & Strategy  1995-2000  
Provo, UT  
Consumer Technologies  Manager- Retail Stores  1989-1995  
Logan, UT  
The Church of Jesus Christ of Latter-day Saints  Volunteer Missionary- Peru  1985-1987  

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
PROFESSIONAL CREDENTIALS

Senior Professional in Human Resources (SPHR)
Teaching Certificate- State of Utah- Secondary Education

PROFESSIONAL AFFILIATIONS

American Society for Training & Development
Society for Human Resource Management
Academy of Management
Organization Development Network
ABSTRACT

Multi-rater feedback, or 360-degree feedback, is a commonly used tool for employee development and performance appraisal. While an abundance of research exists around using this feedback tool for both development and appraisal, little research has determined the correlation, if any, between traditional performance appraisal results and developmental multi-rater feedback assessment. As many employee performance development plans include one or more of these elements, it is beneficial to understand how these assessments might overlap or diverge, in order to formulate a complete picture for use in outlining an employee development plan.

The purpose of this study was to identify the relationship developmental multi-rater feedback assessments have to top-down, traditional single-rater performance appraisals. This research looked at Cereal Manufacturing Technicians (N=105) within a food processing facility, comparing results of an annual performance appraisal (conducted by the employee’s supervisor) to the results of that person’s multi-rater assessment.

Results found that mean, median, and mode scores tend to be significantly higher in multi-rater assessments than in performance appraisals. A significant difference in standard deviations was also found between overall scores. Performance evaluations also showed greater variability and range in scores than did multi-rater feedback assessments. Comparison of individual multi-rater scores to individual performance appraisal scores found little, if any correlation.
This study also showed that top performers on individual multi-rater assessments did not generally fall within the highest ranks of performance evaluation scores. Similarly, the lowest performers in one assessment were not necessarily the lowest performers identified by the other assessment. However, top scorers (top 10%) on multi-rater assessments tended to score in the top half of performance evaluation scores. Likewise, those receiving the lowest multi-rater scores (bottom 10%) also tended to be rated within the bottom half of performance evaluation scores.

Finally, this study found that multi-rater assessments and performance appraisals provide different outcomes and are, therefore, not interchangeable. As both provide valuable information, it is recommended that both a multi-rater (developmental) assessment and a performance (evaluative) appraisal be incorporated into an employee’s overall performance and development plan.
Chapter One: Introduction

Chris Argyris (1982) maintains that organization performance is a balance between the individual and the organization levels, and that individual and organizational learning play a key role in this performance. He further emphasizes that learning can occur only when these individuals and organizations are committed to examining assumptions, and are open to feedback. In his article *Teaching Smart People How to Learn*, he goes on to state that leaders must “reflect continually on their own behavior, identify the ways they often inadvertently contribute to the organization’s problems, and then change how they act.” (Argyris, 1991, p. 100)

In their book *The Extraordinary Leader*, Jack Zenger and Joe Folkman (2002) support this concept in stating:

The Learning and development process relies heavily on feedback to sustain it. There is often a huge gap between how leaders see themselves and how subordinates perceive them. The best way to close that gap is with feedback. Enormous barriers to feedback exist inside organizations. The good news is that when I am presented with disconfirming information about myself, I will more often change how I act rather than change how I see myself. That is the great power of feedback. (p. 241)

Many employee development programs take advantage of some form of structured feedback. Within a number of well-respected leadership development organizations, such as the Center for Creative Leadership, feedback is at the core of their development efforts. Multi-rater feedback is being used increasingly within
organizations as a methodology for both employee development and the attempted measurement of leadership style. Based on the philosophy that individuals should receive a full “360-degree” picture of performance by gaining multiple perspectives, multi-rater feedback gathers input about an individual’s performance by soliciting feedback from those stakeholders impacted by that individual. Multi-rater feedback, or 360-degree feedback, refers to feedback from a variety of people within that person’s circle of influence. Similar to the 360 degrees of a circle, an individual is figuratively at the center of that circle, and feedback is gathered from those in positions to observe the person’s performance: supervisor, peers, customers, direct reports, etc.

The use of formal multi-rater feedback within organizations has increased dramatically over the last two decades (Church & Bracken, 1997). This feedback has become a significant portion of management development expenditures in many organizations. Multi-rater feedback use has steadily increased in popularity, due largely to the internet and easing up of administrative challenges associated with paper-based surveys (Atkins & Wood, 2002). Some studies suggest that over one-third of U.S. companies use some type of multisource feedback process for managers (Bracken, Timmereck, & Church, 2001). Others claim that this estimate is closer to 90% of all Fortune 500 firms (Edwards & Ewen, 1996). Regardless, according to Edwards and Ewen, the question is no longer whether companies use 360-degree feedback, but how.

Numerous studies on the use of multi-rater feedback (that will be presented throughout this document) have raised questions about the effectiveness of these assessments for purposes other than strictly for employee development. Research and opinions are mixed as to where multi-rater assessments and traditional performance
appraisals fit, overlap, and belong in an employee development plan. There are two clear camps on the issue of whether multi-rater assessments should be used for appraisal purposes, or focused exclusively on employee development. There are also a number of opinions to be found somewhere between the two extremes. However, there exists very little published research that addresses the correlation between multi-rater assessments used for development purposes and a traditional single-rater performance evaluation conducted by a supervisor.

Feedback for performance appraisals and feedback for developmental purposes are also often confused or combined. The primary difference between these two is that performance appraisals provide feedback on what is done. Developmental feedback generally provides information as to how the job is done (London & Beatty, 1993). Appraisals look at the past, with the primary intent being an evaluation of historical performance. Developmental feedback, on the other hand, holds as a primary objective the effectiveness of future performance by looking at past behavior.

Despite the increased use of multi-rater feedback as an addition to or replacement for performance appraisal, there is little research regarding the correlation between multi-rater feedback and traditional top-down performance evaluations. If the purpose of a top-down performance appraisal is to measure what was done, and the purpose of a developmental multi-rater assessment is to measure how it was done (London & Beatty, 1993), theoretically each could yield different results. Practically speaking, we may see managers who are able to get results (the what), yet the manner in which they go about attaining these results (the how) is ineffective or inappropriate. As an example, a manager who continually meets performance metric criteria (such as financial or
operational targets) may also experience poor team or work unit morale and high turnover due to the manner in which he or she goes about achieving these targets. Conversely, those who exhibit leadership behaviors as defined through developmental multi-rater feedback may not necessarily reach set financial or operational targets.

Published research has not shown whether those people receiving higher scores in multi-rater assessments (which typically measure behavior) are also those people receiving corresponding high scores on top-down performance appraisals (which typically measure performance). Taking this one step further, it is important to understand when planning for the development of an individual employee, which of these forms of assessment should be taken into account in designing the plan.

Problem Statement

Many organizations are making significant financial investments in employee development. Developmental multi-rater feedback often plays a major role in employee development plans, as does traditional top-down, single-rater performance evaluation. However, these two methods of formal feedback are generally designed with different intended outcomes, and often measure very different aspects of human performance and behavior. If developmental multi-rater feedback and single-rater performance evaluations truly measure different outcomes, they should not be used interchangeably in performance development plans. Conversely, if both forms of evaluation demonstrate similar end results, organizations should not devote the time and expense associated with using both types of assessments in their performance development process. Understanding the correlation, if any, between the two types of assessments would
provide a better understanding of how these two components should be used jointly, separately, or not at all in developing employees within an organization.

**Purpose of the Study**

The purpose of this study is to identify the relationship developmental multi-rater feedback assessments have to top-down, traditional single-rater performance appraisals. The study will examine whether results of the former directly parallel those of the latter. The study will also attempt to identify patterns where results between the two forms of assessment are in conflict, if this should be the case. Specifically, this study will look at Cereal Manufacturing Technicians within a large division of a food processing facility, comparing results of an annual performance appraisal (conducted by the employee’s supervisor) to the results of that person’s multi-rater assessment.

**Research Questions**

This study will address the following two questions:

1. What correlation exists, if any, between the developmental feedback obtained via multi-rater assessments and single-rater, top-down performance appraisals?

2. Do those receiving high top-down, single-rater performance evaluation scores also receive high multi-rater scores (feedback results)? Conversely, do those receiving low single-rater performance evaluation scores also receive low multi-rater scores (feedback results)?
Significance of this Study

The use of multi-rater feedback continues to increase in organizations throughout the world (Church & Bracken, 1997). It is essential that this increased use of multi-rater feedback be met with a commensurate return on investment. A continuing debate exists on whether multi-rater feedback assessments should be used purely for development, or if they are appropriate for performance appraisal purposes as well. While answering this question is not the ultimate intent of this study, it is possible to understand what relationship multi-rater feedback results have to performance appraisal results. In turn, this understanding will serve as a basis for determining whether each draws similar results, and whether one, both, or neither form of assessment should be included in an employee's overall performance development plan.

While an abundance of research exists around using multi-rater feedback for both employee development and employee appraisal (London & Beatty, 1993), it appears that little research has determined the correlation, if any, between traditional performance appraisal scores and multi-rater feedback assessment. As many employee performance development plans include one or more of these elements, it would be beneficial to understand how these assessments might overlap or diverge, in order to formulate a more complete picture of an employee that can be used to construct a development plan.

Conceptual Approach

The conceptual framework for this research is based primary in three areas of literature: (a) the emergence of multi-rater feedback as a tool used in employee development, (b) the use of multi-rater feedback for employee appraisal purposes, and (c)
the major differences between multi-rater developmental feedback and traditional top-down single-rater performance appraisals.

Definitions and Terms

The following are key terms used throughout this study:

Assessment. An evaluation of performance, behavior, or ability. For purposes of this manuscript, the term “assessment” will include both performance appraisals and multi-rater feedback.

Developmental feedback. Information that addresses past behavior or results, presented with the primary intent to provide insight into how an employee might grow personally or professionally. Unlike a performance appraisal, developmental feedback does not place primary focus on evaluating past performance (Fleenor & Prince, 1997).

Employee development plan. A plan prepared by an employee, generally in conjunction with his or her supervisor, intended to map out a course of career progression for the future. This plan often outlines developmental opportunities intended to provide the employee with knowledge, skills, experience, and resources to reach the defined goal.

Feedback. Information that is provided in order to address past behavior or the results of that behavior, to either reinforce or suggest changes in performance or behavior (Zenger & Folkman, 2002).

Multi-rater feedback. Feedback provided by multiple people about an employee’s performance or behavior. Multi-rater feedback is also commonly known as 360-degree feedback, multi-rater assessment, multisource assessment, and upward feedback (Fleenor & Prince, 1997).
Performance appraisal. A formal, written evaluation of an employee’s performance over a period of time, generally completed by the employee’s supervisor. (Campbell & Lee, 1988). A performance appraisal is also known as a performance evaluation, as used herein. As these types of evaluations are historically typical, they are also referred to in this manuscript as “traditional single-rater appraisals.”
Multi-rater feedback is used extensively today for both performance appraisal purposes and developmental purposes. However, as indicated in the previous chapter, opinions are divided as to where multi-rater feedback best fits as a component of an employee’s overall performance development plan.

To more completely understand these issues, it is necessary to review the literature on multi-rater feedback in more depth. This chapter will look at the role of employee development in organizations, the role and immergence of formal feedback programs, and the uses of multi-rater assessment in formalized feedback processes within organizations.

The Need for Individual Employee Development

Over the past 2 decades, the non-book value of all companies rose from 5% to 72%, and is still accelerating. Non-book value takes into account many of those assets which have traditionally not been measured, including organization know-how, management teams, processes, brand recognition, client relationships, and systems infrastructure (Libert, Samek, & Boulton, 2000). With the global economic shift from mass production and manufacturing to one in which knowledge is at the forefront, success increasingly depends on the ability to identify, understand, develop, and leverage the talents, knowledge, and skills that exist within an organization. Yet, as Edgar Schein (1996) identifies, a key theme in many occupational cultures remains the preoccupation with establishing processes that design humans out of the systems rather than into them.
The increased focus on intellectual capital has expanded the need for employee development. Where we once saw organizations investing primarily in physical assets, we now see a shift toward the development of intellectual assets: including the development of employees. Despite this increased emphasis, human resources guru Jac Fitz-enz (2000) states, "Paradoxically, employee development is one of the most important issues of the foreseeable future and one of the worst managed. In fact, it would be an overstatement to claim that it was badly managed. Our experience is that it is unmanaged." (p. 124)

Organizations are investing a significant amount of financial resources, time, and effort in the planning and implementation of employee development programs; many would argue that it is not enough. According to The American Society for Training and Development (ASTD), data collected in 2004 indicates that direct costs for training and development by U.S. firms typically amount to 2% of payroll, while indirect costs and opportunity costs may raise the total to 10% or more. Of this expenditure, an average of 28% of learning budgets went to management development, including first-line supervisors, middle managers, senior managers, and executives. An average of 10% of learning budgets was allocated to executives and senior managers (Sugrue & Kim, 2004).

According to ASTD, average annual employee development expenditure per employee in large organizations averaged $1,190 US in 2004 (Sugrue & Kim, 2004). These investments, however, are not without return. ASTD found that one standard deviation increase (equal to an average of $680 US) in a firm’s annual per employee development expenditure generates, on average, a six percentage point improvement in total shareholder return (TSR) the following year (TSR includes change in stock price.
and any dividends issued during that year). Similar results were found for gross profit margin, return on assets, income per employee, and ratio of price to book value (Bassi, Ludwig, McMurrer, & Van Buren, 2000).

It is clear that investing in employee development is critical to the success of an organization. It is equally important that this investment generate commensurate, measurable returns.

The Importance of Feedback to Development

Much as gauging performance measures was expected in mass production and manufacturing environments of the past, the concept of measuring the effectiveness of these “intellectual assets” has been extended into the knowledge-based arena. Formal individual feedback processes were put in place to meet this demand for measurement of employee performance, and are generally utilized for two primary purposes: performance appraisals and individual development.

According to Van Velsor, McCauley, and Moxley (1998):

The best developmental experiences are rich in assessment data. Assessment is important because it gives people an understanding of where they are now: what their current strengths are, the level of their current performance or leadership effectiveness, and what are seen as primary development needs. In the context of their everyday work, people may not be aware of the degree to which their usual behaviors or actions are effective; in the face of a new challenge, they may not know what to continue doing and what to change. Even if they do realize that what they are doing is ineffective, people may believe the answer is to just work
harder; it may not occur to them to try a new strategy. But when an experience provides feedback on how they are doing and how they might improve, or provides other means for critical self-reflection, the people are more likely to understand their situation and to capitalize on a learning opportunity." (p. 9)

According to Seashore, Seashore, and Weinberg (2001), feedback is information about past behavior, delivered in the present, which may influence future behavior. The term “feedback” is taken from the field of cybernetics, and originally relates to the theory of control (Seashore et al., 2001). It deals with the relationship between two or more interacting systems. Norbert Weiner, the father of cybernetics in North America, pioneered the use of feedback mechanisms to help the military improve accuracy of fire on moving targets. Kurt Lewin took Weiner’s concepts from cybernetics and applied them to human interaction.

Peter Senge (1990), in his work The Fifth Discipline, describes the concept of feedback as the basis for systems theory. Feedback is critical for survival. Without feedback, a hand would likely not be withdrawn from hot water or a flame. Without feedback, people hallucinate and withdraw. Feedback is life: to live means to interact with the environment (Seashore et al., 2001).

In Kurt Lewin’s model of change, he articulates a process which involves steps of unfreezing present individual and organizational behavior, moving towards the change, and refreezing. Linda Ackerman-Anderson (2003) further defines three levels of change: Developmental, Transitional, and Transformational. The most involved of these, transformational change, according to Ackerman-Anderson, requires “wake-up calls” or
mindset shifts. These wake-up calls come as the result of feedback. The ability to change, and driving impetus behind unfreezing present behavior, is based on feedback.

Many organizations tout the value of individual feedback as a form of both measurement and development, but may not fully realize the potential impact formal feedback mechanisms can have. On the organization level, feedback may call attention to important performance dimensions and organization climate that impact the financial outcome of an organization (London & Beatty, 1993). According to Daniel Goleman (2000):

Managers often fail to appreciate how profoundly the organizational climate can influence financial results. It can account for nearly a third of financial performance. Organizational climate, in turn, is influenced by leadership style—by the way that managers motivate direct reports, gather and use information, make decisions, manage change initiatives, and handle crisis. (p. 21)

Robbins (1990) states that systems—organizations or individuals—continually receive information from their environment which helps the system adjust and take action in order to correct deviations from a defined course. In his widely recognized work *The Fifth Discipline*, Senge (1990) claims that this ability to question and reflect upon the current state through feedback is a key element in the makeup of a “learning organization.” A learning organization, according to Senge, is “an organization that is continually expanding its capacity to create its future. For a learning organization, ‘adaptive learning’ must be joined by ‘generative learning’—learning that enhances our capacity to create.” He further describes systems thinking as the cornerstone of what he refers to as the “Fifth Discipline.” According to Senge, “From the systems perspective,
the human actor is part of the feedback process, not standing apart from it. This represents a profound shift in awareness. It allows us to see how we are continually both influenced by and influencing our reality.” (p. 14)

Early organization theory, however, subscribed to a more scientific and linear view of management and organizations. Classical organization theorists such as Frederick Taylor viewed organizations as closed systems in which variability could be eliminated. The feedback loop was virtually non-existent. Workers were viewed as factors in production. Managers, therefore, had the right to utilize these resources (employees) as they pleased in order to achieve objectives. In this scientific approach to organizations, employees were perceived as logical and rational. When provided with determined input, the same, continuous output would result. The behavioral aspect of management was minimized. The value of personal and organization feedback was largely dismissed.

Organization theory modernists saw basic flaws in classical organization theory. A scientific view of organizations denied the human side of the equation. Selznick (1953) pointed out that organizations and individuals were subject to forces outside of stated, rational goals. The concept of authority and acceptance of authority were also neglected, according to Barnard. Organizations began to see that the human behavioral side played an important role in their success.

Theorists such as March and Simon took this concept even further, suggesting that men (humans) were not “rational.” It could not be determined how they would behave at all times. They asserted that senior managers could not simply and bureaucratically create rules that covered all situations, nor were they capable of
understanding all the variables of which they were likely not aware. In terms of management decision-making, March and Simon (1993) persistently rejected traditional management theory, claiming that traditional theorists did not consider the emergence of goals. March and Simon further introduced the concept of *satisficing*, arguing that employees would not perform to optimize. When employees performed at minimal levels, they were rewarded (fed-back) by receiving a paycheck. The feedback loop consisted of a behavior-reward flow.

General systems theorists, on the other hand, claimed that organizations and individuals function within systems. Understanding the system requires feedback as to how that system acts within its environment, as well as studying each part in interaction with the other parts of the environment. Individuals and organizations, as with everything, according to Ken Wilber (1995), are essentially holons. They are whole in one context, and a part in another. An organ is a whole, composed of cells, yet it is a part of another organism. The holarchy becomes an order of increasing holons: an order of increasing wholeness and integrative capacity (Harman, 1998). Each organization is comprised of individuals (who, in addition to being organically holistic, are experientially, culturally, socially, etc., holons in themselves). These individuals make up subsets of the organization (departments, cultures, etc.) which, in turn make up larger subgroups (regions, divisions, countries, etc.). These larger subgroups assemble to make up corporations, organizations, nations, etc. which, in turn, are holons within a larger global holarchy. However, we often address organizations and individuals as if they were complete entities unto themselves, independent of the elements of the holarchy. Yet these organizations and individuals interact with their environment. They both impact
and are impacted by the environment around them. This interaction often determines the degree of success or failure.

Chris Argyris (1982) maintains that learning can occur only when people are committed to examining assumptions and are open to feedback. Peter Koestenbaum (2002) identifies two foundations required for leadership as **reflection** and **realism**. He defines reflection as “not just to look but to look at the act of looking itself; not just to think but to think about thinking itself; not just to learn but to learn about learning itself; not just to feel but to examine the act (or passion) of feeling itself” (p. 35). He further states, “Realism means to be in touch with people. You are conscious of how others perceive you. Their perceptions of you need not be accurate; in you, they may see only themselves. But they will act on these perceptions. Understanding that is realism. How others see you—and, even more, how others limit you—these are fundamental realities” (p. 45). He goes on to explain, “Realism means to be aware of how the other person thinks, how others perceive you. That is often much more difficult to bring about than it may appear to be. It is important, for example, to understand how others are affected by the very system in which we live.” (p. 45)

Senge, Kleiner, Roberts, Ross, and Smith (1994) further affirm that self-reflection and inquiry are critical. They state, “Two types of skills are central to this work: They are **reflection** (slowing down our thinking processes to become more aware of how we form our mental models) and **inquiry** (holding conversations where we openly share views and develop knowledge about each other’s assumptions).” (p. 76)

Bolman and Deal (1997), when addressing the relationship between individual leadership and group success, claim that leaders must discover their own spiritual center.
through reflection. According to these authors, as stated in their book *Reframing Organizations*:

Effective leadership is a relationship rooted in community. Successful leaders embody their group’s most precious values and beliefs. Their ability to lead emerges from the strength and sustenance of those around them. It persists and deepens as they learn to use life’s sounds to discover their own spiritual centers. As they conquer their demons within, they achieve the inner peace and bedrock confidence that enable them to inspirit and inspire others. (p. 56)

Koestebaum (2002) further articulates that an individual must be aware of how he or she is perceived by others. He states:

> From the level of a deeper philosophical point of view, we must be aware that a human being, until witnessed by another person, does not know he or she exists. This phenomenon is known as validation. You do not know that you exist until you are reflected in a mirror, and the only mirror to a soul is another soul. (p. 113)

In the 1950s, the team of Joe Luft and Harry Ingham (1955) were researching human personality at the University of California. The result of their research was a model that came to be known as the Johari Window. The Johari Window consisted of a two-by-two matrix, with “self” and “others” serving as the labels for the x-axis and y-axis respectively. The premise on which the model is built indicates that there are aspects of our personalities which are known to us, and aspects that are known to others. Aspects that one may understand about his or herself may not be known to others, and visa versa. The squares within the matrix represented four categories:
The public area: those things which are openly known and openly discussed.

The hidden area: those things that others observe about us, but that we don’t know about ourselves.

The unknown area: that which nobody—ourselves included—knows about.

The private area: aspects of our self that we know about and choose not to disclose to others. (Luft, 1969)

According to Luft and Ingram (1955), knowledge about those areas unfamiliar to us comes only through feedback. We begin to form views that are based on what we think we know about ourselves and others, which in turn dictate the way we view the world (Weick, 1998).

Argyris relates to this concept through the Ladder of Inference (Argyris, Putnam, & Smith, 1985). The actions we take are the cumulative results of our progression up the ladder. We (as individuals and organizations) take in data: the first rung of the ladder. From this observable data and our own experiences, we select the data we find fits our worldview, add meanings, make assumptions, draw conclusions, adopt beliefs, and take action. Any incomplete or distorted step within this process results in improper action.

Carl Jung suggested that the hidden side of our personalities is a frequent source of embarrassment or trouble. He referred to this side as “the shadow”: the envelope or container for the less-developed aspects of our personality. It contains those vaguely familiar parts of our personalities that were largely unknown or declared unwelcome during our upbringing. Karen Horney further suggests that the individual tends to replace
his or her view of real self (which promotes growth) with an inaccurate or ideal self (which limits development). (Segal, 1997)

Understanding ourselves as leaders, and how we interact with others, further helps us to understand what impact we have on others. This is a characteristic essential to leadership. According to Peter Northouse (2001), “The psychodynamic approach places emphasis on leaders obtaining insight into their personality characteristics and understanding the responses of subordinates based on their personalities. Leaders should also encourage work group members to gain insight into their own personalities so that they can understand their reactions to the leader and to each other.” (p. 190)

Several commonalities exist in the literature regarding feedback and the role it plays in individual leadership and success. First, it is clear that our ability to interact with our environment determines, to a large degree, our level of success. Second, our understanding of the impact we have on those around us is dependent upon feedback. Without feedback, our view of reality is created entirely from within our own minds, without external perspective from which to frame this reality. Effective people use both formal and informal feedback mechanisms to formulate this picture of reality. Third, effective organizations take advantage of processes which will encourage this feedback to take place.

Feedback and Performance Management

Feedback, in its various forms, constitutes a major component of what many organizations have come to know as a Performance Development or Performance Management Plan. The Society for Human Resource Management’s (SHRM) 2000
Performance Management Survey showed that performance development planning in organizations is evolving from a system dominated by the performance appraisal to a system that focuses on employee development. However, that process is far from complete (SHRM, 2000). Of those surveyed, over 70% responded that their organizations had written performance development plans for most (at least three out of four) of their executives. Sixty-four percent of the respondents reported that their organizations had performance development plans for most exempt employees, and 45% had plans for most non-exempt employees. Eighty-five percent reported that performance plans were largely done in conjunction with performance evaluations on an annual basis. They further reported that survey respondents were significantly more satisfied with traditional system components of performance management systems (appraisal and discipline) than they were with the developmental components (multi-rater assessment, leadership development, and coaching). According to SHRM, this may be due to the fact that organizations have been using traditional appraisal methods for a longer period, and that organizations have only recently started thinking about development as part of their performance management systems.

According to SHRM, nearly one-third (32%) of survey respondents were dissatisfied or very dissatisfied with their overall performance management systems. This could be largely attributed to the fact that the developmental components of the performance management process were missing. They also cite the lack of support from top management as a major hindrance in the success of the performance management process. In fact, 42% of those survey respondents responded that executives did not review the performance management system at all (p. 9).
According to Diane Marentette (2000) of Personnel Decisions International, in referencing the previous SHRM survey, too many performance development plans are focused solely on measurement of past performance, rather than development. Marentette admonishes that “emphasizing evaluation and expectations rather than personal development puts employees on the defensive and demotivates them to participate actively in the process.” (p. 26)

Traditional employee development plans include four phases: performance planning (including written performance metrics), coaching and feedback, appraisal of past accomplishments, and identifying areas for improvement (Marentette, 2000). In the majority of cases, the appraisal component has become the primary focus. According to Howard Pardue (2000), the increased demand for accountability by government agencies and related legislation have made appraisal systems a necessity in organizations. The Equal Employment Opportunity Commission’s Uniform Selection Guidelines have further forced organizations into implementing appraisal systems. This demand for appraisal is emphasized by the need for a basis from which to make decisions related to promotions, salary, discipline, and termination. In the majority of these cases, it is job performance specifically that is measured. In other words, the evaluation is based on whether or not the person was able to meet the performance standards determined during the performance planning phase. These standards are typically derivatives of quality, speed, and price/costs of measurements.

According to Jac Fitz-enz (1995) in a study of organizational communications, it was found that employees were much more interested in performance and career opportunity information than in any other topic. He claims that there has likely been as
much research and speculation on performance review and appraisals as on any managerial subject (p. 110). Yet performance appraisals often tend to be regarded negatively due to several factors. First of all, needs cannot be entirely met by one performance evaluation. It is improbable, if not impossible, that one performance evaluation would capture all skills necessary to effectively lead in a given organization. Second, appraisals from a single supervisor are inherently problematic, as they provide only one person’s opinion. Success as a leader generally impacts more than just the supervisor, and this success is impacted by the views of more than just the supervisor. This also assumes that the manager providing the feedback is capable of providing appropriate evaluations (Harvey, 1994). Third, these appraisals include limitations that are the result of the following elements: (a) reliance on generalized traits; (b) limited or nonexistent frame of reference for ratings; (c) a memory-based (and often inconsistent) description of past performance; and (d) inaccurate interpretation of the data (Moses, Hollenbeck, & Sorcher, 1993).

Based on review of the literature, it is clear that feedback is a critical part of an effective employee performance plan. However, traditional performance appraisals have often not provided the feedback employees needed in order to improve or develop. Traditional performance appraisal feedback has often been focused exclusively on evaluation of past performance, ignoring developmental feedback entirely. Additionally, feedback from a single supervisor may not provide accurate and comprehensive feedback. Because of this limited and often distorted view of performance, organizations have sought more effective methods for providing formalized feedback to their employees.
The Emergence of Multi-rater Assessment

Edwards and Ewen (1996) assert that no organizational action has a greater ability to motivate employee behavior than feedback from credible work associates. London and Smither (1995) explain that in the socially-constructed world in which many people work today, the judgment of others around them (no matter how biased they may be) constitutes an important reality. Additionally, information about performance is now sought for more than just annual performance evaluation (Fleenor & Prince, 1997).

For the past several decades, the competitive business environment and flattening of organization structures have dictated the need for a better methodology in evaluating employee performance. New business structures encouraged interdependence and reliance upon more than just the few immediate coworkers required previously. The traditional forms of top-down feedback no longer yielded the feedback that was required for success. By gathering feedback from multiple stakeholders, multi-rater assessments provided a full-circle view of performance.

Multi-rater feedback today is commonly known as 360-degree feedback. The term 360-degree feedback was originally trademarked by TEAMS International (Bracken et al., 2001). Despite the trademark, the term 360-degree feedback has been used commonly for the past two decades when referring to this type of assessment. Among others, common terms for this type of feedback include: multi-rater assessment, multi-rater feedback, multisource assessment, and upward feedback.

Multi-rater feedback has recently gained even greater popularity. Wingrove (2001) suggests several possible reasons for the recent boom in the use of multi-rater assessment:
1. Contemporary technology now makes it easier to deliver.

2. Human resources, as a profession, has become much more data thirsty. They are increasingly being expected to provide concrete data as to employee performance, as well as their own performance as a human resources function.

3. The timing was right. Many of the other more recent performance measurement "flavors of the month" have not had the mass appeal that total quality, process reengineering, and balanced scorecards once did. Yet, the quest for a compelling initiative is never-ending.

Although extensive use of multi-rater feedback prior to the 1980s was rare, the concepts behind multi-rater feedback are not new. Assessment centers developed by the German Military during World War II recognized the value of receiving feedback from multiple sources (Fleenor & Prince, 1997). Also during this time period, various others explored the use of a form of multi-rater feedback via the technology of T-groups.

One of the earliest recorded uses of surveys to gather information about employees occurred in the 1950s at Esso Research and Engineering Company (Bracken, Dalton, Jako, McCauley, & Pollman, 1997). During the 1960s and 1970s, a great deal of interest arose around how best to provide employees with a broader, more accurate view of their performance. Researchers began to investigate the use of feedback from not only an employee’s supervisor, but from his or her direct reports as well. A number of companies began to incorporate feedback from direct reports into their performance discussions.

In the mid-1980s, the Center for Creative Leadership began conducting studies on the use of multi-rater feedback. Three key findings of the study drew attention to the
potential value of 360-degree feedback. The first was that individual feedback is an important potential growth element for both professional and personal development. Second, it found that most successful leaders were learners; they made everything they did a learning experience. When these leaders received new information, such as through feedback, they analyzed the information and adjusted accordingly. Third, they found that most people, especially in the middle and senior manager ranks, operated in a feedback-poor environment (Lepsinger & Lucia, 1997). Other studies have also supported these findings (Marchese & McGowan, 1995).

Organizations and individuals began to see that multi-rater feedback would allow them to identify both strengths and development needs that might not be exposed in traditional performance evaluations (Jones & Bearley, 1996). According to Fleenor and Prince (1997), in supporting claims made by Jones and Bearley (1996), there were three organization shifts that made the use of multi-rater feedback all the more necessary: (a) the change in focus from management skills to leadership skills, (b) the change in dependency to self-responsibility in career planning, and (c) the move away from traditional hierarchy and structure. It became apparent that traditional performance appraisal was less than adequate in employee development and career planning. In traditional appraisal, supervisory ratings were generally the sole source of evaluative data. However, subordinates are often better positioned to view and evaluate leadership behavior. As it is the subordinate who is most affected by the manager’s leadership behavior, receiving feedback from that subordinate would offer a more complete view than the sole opinion of the supervisor. Additionally, the supervisor will not likely observe the ratee’s leadership abilities on a day-to-day basis. People who relate to a
manager at different times and for different reasons may perceive the manager differently (London & Beatty, 1993).

Many managers have not received as much honest feedback as is necessary for an accurate self-perception (Waldman, Atwater, & Antonioni, 1998). One of the inherent advantages of multi-rater assessments lies in the ability to compare one's self-perception with that of other raters. Comparing self-views with the views of others promotes a self-awareness that can lead to an "unfreezing" process in which a manager is motivated to rethink his or her behavior and its impact on others. In a 1996 study published in Organizational Dynamics, it is reported that 25% of the feedback received on multi-rater assessments was expected positive feedback, while 30% was unexpected positive feedback. On the other side, 20 to 30% was expected negative feedback, and 15 to 20% was unexpected negative feedback. The study also showed that nearly 20% of those receiving feedback would be surprised by their low ratings (Antonioni, 1996).

As indicated by the literature, the lack of agreement between self-ratings and the ratings of others is one of the key advantages of multi-rater feedback. These gaps in perception, as earlier identified in this manuscript, provide insight into how close or how distant our awareness of performance matches that of those around us.

Variance in Rater Perception

According to Seashore, Seashore, and Weinberg (2001), in order to believe feedback—to fit it into our current pattern of thoughts, beliefs, and values—we must be able to overlap that feedback with what we already know. The more information contained in the feedback, the less likely it will overlap with what we already "know."
This requires more rethinking, reorganizing, revaluing, and replacing it with that which we already “know” to be true.

Geoff Bellman (1996) also suggests that often the feedback we receive may not match our own opinion. He refers to this type of feedback as “lighting the shadows” (p. 49). We then face the dilemma of being confronted with “two truths” (p. 72). When what we already “know” is radically different from what the evidence (feedback) provides, we pick one “truth” over the other and live with the consequences.

The difference in perception between rater groups is a useful source of information not available in top-down, single-rater appraisal. Discrepancies between self-ratings and feedback from others can also lead feedback recipients to perceive a need to change their behavior (London & Smither, 1995). The weight of numbers often forces managers to reconsider their own impressions of performance and behavior (London & Beatty, 1993).

According to a 1994 *Wall Street Journal* article (Lublin, 1994), it is difficult for many managers to accept feedback, especially from subordinates. One of the advantages of multi-rater assessment is the unique opportunity for individuals to rate themselves, and compare these ratings to others’ ratings of their skills and behavior (Fleenor & Prince, 1997; Hazucha, Hezlett, & Schneider, 1993; London & Beatty, 1993; Smither, London, Vasilopoulos, Reilly, & Salvemini, 1995; Yammarino & Atwater, 1993). Studies examining discrepancy between self-assessments and others (supervisors, peers, direct reports, and customers) indicated that the ratings between self-scores and scores provided by others showed significant disagreement (Lublin, 1994). Self-ratings tend to be inaccurate when compared to the more objective ratings of others (Yammarino &
Atwater, 1993), with self-ratings, in general, being significantly higher than the ratings of others (Nowack, 1992). However, these studies also argue that it is this very difference in perception that prompts the increased self-awareness that leads to change.

Studies conducted by Farh, Dobbins, and Cheng (1991) indicate that the phenomenon of self-leniency (the tendency to rate oneself higher than others’ ratings) is a cultural occurrence unique to Western tradition. In reviewing the use of multi-rater assessments in Taiwan, they found the opposite to occur. A modesty bias was found, in that Taiwanese participants typically rated themselves lower than the ratings given by their raters. However, in subsequent attempts to replicate the study with Chinese managers, no modesty bias was found (Yu & Murphy, 1993).

According to Ostroff, Atwater, and Feinberg (2004) in citing several studies (Ashforth & Mael, 1989; London & Smither, 1995), “Theorists have clearly made the point that even if others’ ratings are not objective or accurate, it is important for individuals to understand how they are perceived by others in order to navigate the political realities in organizations” (p. 343). Fletcher (1997) suggests that if self-perceptions differ from those of others “then it is difficult to see how one can manage work relationships successfully, contribute well as a team member and adapt one’s behavior to circumstances and individuals” (p. 186). Self-other agreement has been associated with better performance evaluations, promotions, or other measures of career success (Bass & Yammarino, 1991; Church, 1997). However, many ratees find that others’ perceptions of them are not how they view themselves. People who are highly vested in their own self-image may have trouble reconciling performance discrepancies (Kenny & DePaulo, 1993). Nowack (1992) maintains that inflated self-assessments may
also interfere with an individual’s development, as they may not recognize the need for improvement. Those ratees who have the ability to monitor and change their own behaviors will be more accustomed to reacting to feedback from others (Gangestad & Snyder, 2000).

This discrepancy in ratings is the subject of a great deal of research. In fact, a significant portion of the literature on multi-rater feedback has focused on examining self-observer rating discrepancies (Atkins & Wood, 2002; Atwater, Roush, & Fischthal, 1995; Bass & Yammarino, 1991; Church, 1997; Fleenor, McCauley, & Brutus, 1996; Nilsen & Campbell, 1993; Ostroff et al., 2004), or on changes of ratings over time (Reilly, Smither, & Vasilopoulos, 1996; Smither & Walker, 2001; Smither, Wohlers, & London, 1995). In making a compilation of the research, Tornow (1993) notes that there are three core assumptions that relate to the discrepancy between self-ratings and the ratings of others: (a) Awareness of how one’s skills and behaviors are perceived by others is a key to effectiveness as a manager, (b) managers need to improve in the area of self-objectivity, and (c) these differences in self-other perceptions motivate managers to alter their self-views and/or improve performance. However, this variance continues to be a source of controversy amongst assessment purists in terms of the validity and reliability of multi-rater assessments in general.

Research on rater congruence can be broken down into three main areas of focus: self-other congruence, other rater congruence (also referred to as other-other congruence), and within-source agreement (Kustis, 2001). In a study of naval academy student leaders, correlations of $r = -.09$ were found for self-supervisor ratings, $r = .19$ for self-subordinate, and $r = .35$ for supervisor-subordinate correlations (Atwater &
Yammarino, 1992). Similar results were found by Carless, Mann, and Wearing (1998), with $r = .15$ for self-supervisor, $r = .07$ for self-subordinate, and $r = .35$ for supervisor-subordinate relationships. They also found agreement ($r = .46$) between pairs of subordinates which, according to the researchers, indicates the multi-rater questionnaires measured stable features of behavior. A meta-analysis conducted by Harris and Schaubroeck (1988) reported a supervisor-peer correlation of $r = .62$ (Atkins & Wood, 2002). Atkins and Wood further report that there “appears to be some consensus that others’ ratings are more closely related to one another than they are to self-ratings” (p. 873). Additional studies indicate that those whose self-evaluations on multi-rater feedback most closely matched the opinion of their raters were also likely to receive better performance evaluations, promotions, or other measures of success (Bass & Yammarino, 1991; Church, 1997; Ostroff et al., 2004).

In addition to the ongoing debate regarding rater discrepancy, there is further discussion centered on the appropriateness of the most commonly-used rating scale, the Likert scale. The vast majority of multi-rater assessments use some variation of the Likert scale. Five-point scales and seven-point scales are common. Numerous researchers have their opinions on just how many points should be included in the scale. However, Human Resources firm Development Dimensions International (DDI) questions the effectiveness of using the Likert scale for multi-rater assessments altogether.

Jennifer Salopek (2004) states, “What Likert did not anticipate was the difficulty the average person encounters in using a 5- or 7-point scale for assessment. Effectively using the scale calls for a degree of nuance that many raters may not understand” (p. 28).
She further explains that, in using a Likert rating scale, raters may have to work very hard at determining what they should work on. Raters, as well as ratees, may have a difficult time understanding the ambiguous differences between a rating of 3.6 in one area and a 4.0 in another. This, in part, contributes to what DDI claims to be the most difficult part of multi-rater assessment: Only 1 in 10 participants makes any change in his or her behavior after receiving a multi-rater assessment (Salopek, 2004).

Logic would seem to dictate that multi-rater assessment results would improve when looking at the same subject group over a period of time. Several studies (Hazucha et al., 1993; London & Wohlers, 1991; Walker & Smither, 1999) appear to support this logic. However, as indicated by Atkins and Wood (2002), these studies cannot be used to validate the accuracy of the multi-rater process, as it is impossible to isolate the locus of any improvement. In fact, Siefert, Yukl, and McDonald (2003) propose that there is little evidence that the multi-rater process results in change. Smither, London, and Reilly (2005) claim that this is due to the fact that most feedback programs suggest participants focus on only a few key areas for change. Subsequently, the recipient may only make meaningful changes in those areas; yet those changes may be significant in overall on-the-job performance. These changes, however significant they may be, would have little impact on the average ratings.

Although few longitudinal studies can be found, several studies contradict the view that multi-rater assessment processes do not result in change. In a Walker and Smither (1999) study over 5 years involving 252 managers, no improvement in overall ratings was found between the 1st and 2nd year, but higher scores were noted between 2nd and 3rd and 3rd and 4th yearly administrations. A study of 92 managers over 30
months by Reilly et al. (1996) found that performance (as defined by the average ratings by subordinates) increased. Managers whose initial levels of performance were low improved between the 1st and 2nd administrations, and sustained this improvement 2 years later. The researchers noted that this performance improvement could not be attributed solely to regression to the mean, and was not related to the number of times, or when, managers received feedback. These results suggested that the continued administration of an upward feedback program can result in sustained change over time. Despite studies that appear to indicate multi-rater assessment results may improve over time, there exists little published research to support the notion that on-the-job performance improves over time as a direct or indirect result of multi-rater assessment.

Lublin (1994) asserts that many critics of upward feedback feel that it is difficult to collect accurate and valid ratings. McGarvey and Smith (1993) claim that the effects of recency—the influence of recent events on survey outcomes—may also distort the validity of the instrument. Vinson (1996) makes the claim that multi-rater assessments often generate conflicting opinions, and that there may be no way to determine whose feedback is accurate. Results are also skewed by the stated and intended purpose of the assessment, as will be detailed later in this manuscript.

As discussed previously, one of the reasons for discrepancy in rater scores is that rater groups observe ratees from differing perspectives and in varying situations (London & Beatty, 1993). This discrepancy in perception among rater groups begs the question as to whether or not raters in similar rating classes (i.e., peers compared to peers) rate performance differently because they have different views of what constitutes good performance in a given role (Smither et al., 2005). A disparity in views would certainly
result in a variation in the ratings that raters would provide. However, studies show that raters in similar roles, as well as in different roles, shared a common view of performance dimensions (Facteau & Craig, 2001; Maurer, Raju, & Collins, 1998; Scullen, Mount, & Judge, 2003; Smither et al., 2005).

Jones and Bearley (1996) claim that no multi-rater instrument is inherently valid or invalid, and that the validity of the instrument is always situation-specific; it does not reside in the assessment instrument. They further state that the multi-rater process itself—particularly the follow-up—is what generates the payoff. Atkins and Wood (2002) argue that:

[The] literature has had little to say about validity because studies in this tradition have inevitably confounded issues of validity with other theoretical issues. For example, a reduced correlation between two sources of ratings might reflect poor reliability, poor validity, or quite valid differences of opinion based on having witnessed different behavior of the feedback recipient. (p. 873)

Murphy and Cleveland (1995) note that assessment ratings cannot be viewed as pure measurements of performance. Performance ratings reflect a complex interaction between what the rater is doing, the goals and purposes of the rater, and the context in which the ratings occur. Murphy, Cleveland, and Mohler (2001) state:

If we define validity in terms of the question ‘do these ratings reflect the person’s true performance level?’ we are likely to come to different conclusions about validity than if we ask another set of questions (‘Do these ratings reflect other people’s perceptions of performance and effectiveness?’). That is, multisource ratings may or may not yield information about ‘true performance’; though there...
is evidence for the construct validity of performance ratings, there is also clear evidence that factors other than performance influence ratings, regardless of the source.

Multisource ratings are more likely to provide valid information about how one’s performance is perceived at various levels of the organization, and regardless of whether these perceptions are accurate, it should be useful information to find out that your subordinates, peers, and so on believe that you are effective or ineffective in various aspects of your job. (p. 138)

Walter Tornow (1993), in an article published in *Human Resources Management*, examines whether a multi-rater assessment is an end in itself (the scientific perspective) or a means to an end (the practical perspective). From a scientific perspective, the goal would be to maximize the validity of the instrument by reducing error and rater variation. On the other hand, the practitioner values the differing input of the rater variation as information from which much can be learned. Turnow goes on to argue that managers in an organization are practitioners rather than scientists, and that the feedback process is an insightful and valuable means to an end. When multi-rater processes use these discrepancies in rater perspectives as feedback in itself, they are powerful tools for development.

*Multi-rater Assessments for Development vs. Appraisal*

For many organizations, multi-rater feedback has become a perceived panacea for assessment difficulties. This increased popularity has also distorted the original
developmental purposes of multi-rater feedback. Multi-source feedback systems were originally intended for developmental purposes (Waldman et al., 1998). However, they are increasingly being used for appraisal purposes. DeNisi and Kluger (2000), in adding to the research of London and Smither (1995), report that 85% of multi-rater feedback systems are used for development (with 50% being used exclusively for this purpose). A 1996 poll of 750 American Management Association member companies, 90% of the respondents with multi-rater systems use the results for both development and evaluation (Bohl, 1996). Although performance appraisal may have a developmental component in that it seeks to improve weaknesses and enhance strengths, according to London and Beatty (1993) this aspect is often incidental, as the focus of managerial evaluation is on the performance of the work unit as a whole. As supervisors are accountable for their subordinate’s work units, they may be less concerned with leadership behaviors that they may seldom observe.

A great deal of controversy exists as to whether multi-rater feedback should be used for performance appraisal purposes, or exclusively for development. Research has shown that employees believe appraisals used for development are more likely to produce positive outcomes than appraisals used for administrative purposes (Bettenhausen & Fedor, 1997). In follow-up studies carried out after multi-rater feedback assessments were conducted, 34% of the subordinates surveyed said they would have rated their boss differently had the feedback been used for the manager’s performance appraisal (London, Wohlers, & Gallagher, 1990).

Arguments have been made both for and against the use of multi-rater feedback for appraisal. Those in favor of using the process for administrative purposes (e.g.,
appraisal, pay, promotion, etc.) are quick to point out that peers and subordinates are critical sources of information (Dalton, 1996). Excluding feedback from these groups would lessen their ability to understand and, consequently, improve upon their own performance. Additional arguments suggest that without the accountability associated with using multi-rater feedback as a performance appraisal, there is not incentive for the manager to change (Dalton, 1996; Kustis, 2001).

Although multisource feedback systems were originally used almost exclusively for developmental purposes (London & Smither, 1995), the trend is for organizations to use these for administrative purposes (Dalessio, 1998; London, 2001). Fleenor and Prince (1997) suggest that these trends are due, in part, to changes in the business environment. Employees who once felt satisfied and secure in their jobs now find themselves responsible for their own careers. Because of this, viewpoints from multiple perspectives are critical to overall success. They further explain that downsizing has made it important for even skilled managers to be aware of their own strengths and weaknesses in order to remain competitive.

Some organizations currently integrate multi-rater feedback results with performance evaluation, although most continue to use multi-rater feedback mainly for development (London & Beatty, 1993). Many organizations are feeling the pressure to make feedback that is intended for development more evaluative, linking it to a manager’s or employee’s performance appraisal. This impetus comes largely from the desire for organizations to “get their money’s worth” out of their developmental assessments (Waldman et al., 1998). Despite these trends, most researchers are of the
opinion that multi-rater feedback should be used for development purposes only (Ghorpade, 2000).

During the past several decades, traditional top-down, evaluative performance appraisals have been criticized. According to Harvey (1994), a number of factors contribute to the negative feelings surrounding performance appraisals, including: (a) One appraisal does not fulfill all needs; (b) a top-down, single-source appraisal is inherently problematic; (c) developmental feedback tends to be lost or greatly overshadowed by the performance appraisal aspect; and (d) effective performance evaluation requires skills many managers lack. The basic premise behind using multi-rater assessment for appraisal and administration was derived out of concerns similar to these factors, as well as the philosophy that appraisal could benefit from the advantages of multiple raters (as had been the case for developmental feedback). In fact, when confronted with a completed performance appraisal, self appraisals generally disagreed significantly with supervisor appraisals (Harris & Schaubroeck, 1988). This also created legal challenges, in that the feedback may have been received from a single supervisor who was, as in many cases, inadequately skilled in performance evaluation. Consequently, it was perceived that feedback from multiple sources would mitigate part of this issue.

The advantage of soliciting multiple perspectives is also a disadvantage in using multi-rater feedback for appraisal. Feedback from multiple perspectives takes into account the views of multiple people. However, in so doing, it adds a layer of administrative complexity that did not exist in top-down, single-rater appraisal (London & Beatty, 1993).
Harvey (1994) suggests that organizations should: (a) replace single-source, top-down assessment information with data obtained from multiple raters; (b) separate evaluation for salary administration and compensation from assessment designed to help individuals develop and improve; and (c) ensure that performance appraisal improvements are received as positive changes.

There is growing criticism regarding the belief that performance appraisals do not address those elements referred to by Lepsinger and Lucia (1997) as "style" that are so critical to success. These authors define "style" as patterns or ways of relating to the external environment (e.g., self-confidence, energy level, self-sufficiency, or emotional stability). However, traditional performance appraisals are generally limited to what most would consider concrete, measurable behaviors. Yet it is those personal qualities often presumed to be subjective measures that differentiate the outstanding employee from the mediocre employee (Collins, 2001; Goleman, 1995, 2000; Goodson, 2000; Lepsinger & Lucia, 1997; Zenger & Folkman, 2002). Ward (1997) claims that we are now seeing an increasing tendency to want to assess people by looking at not only the results they have achieved over time, but also how they were achieved.

Traditional performance appraisals and developmental multi-rater assessment share several commonalities. Both involve reports of behavior, as well as judgments of performance based on results. Both can be subject to traditional issues of data validity and bias, for example, response consistency, leniency, halo, and stereotyping (London & Beatty, 1993). However, there are also distinct differences between the two uses.

Performance appraisals are conducted for evaluative purposes and have organizational consequences, such as pay, promotion, ranking, or even disciplinary
action. Performance appraisal is not ordinarily geared to improving work unit performance or leader behavior (London & Beatty, 1993). The purpose of developmental multi-rater feedback, on the other hand, is to measure behavior rather than job-specific performance. Developmental feedback measures how things were done, versus what was done (Harvey, 1994). This generally makes appraisal less subjective, as it often relates to hard measurements. Appraisal is often backward-looking and historical, measuring whether or not expectations were met. Developmental feedback is generally conducted with the intent of improving future behavior.

Another major distinction between these two uses is ownership. When feedback is used for appraisal purposes, the feedback is often not private, and "owned" by the rater or supervisor. However, the work of Herb Meyer and others at General Electric as early as 1965 suggest that public, quantitative performance appraisals do little to produce change (Dalton, 1996; Meyer, Kay, & French Jr., 1965). When used for developmental purposes, the data are private, or owned by the ratee. In fact, in many cases, developmental multi-rater feedback is not shared with the ratee's supervisor, although he or she may be required to work with the supervisor on a development plan based on these results. According to Edwards and Ewen (1996), this confidentiality is key. They claim that the credibility of the multi-rater process depends on two main elements: the anonymity of those providing the feedback and the confidentiality of the feedback itself.

Preserving the integrity of a purely developmental exercise also makes it less threatening for peers and subordinates who might have reason to fear retaliation (London et al., 1990). However, Dalton (1996) points out that anonymity is no longer a guarantee against reprisal if all raters in a particular group give the manager a very low score.
Despite this caveat, providers of feedback may be more willing to be completely honest, as the perception is established that neither they nor the recipient have anything to lose. Receivers of this feedback may also be less defensive, more open to others’ views, and more apt to regard the feedback as credible. Crucial to this process is who has access to the results.

Numerous studies have concluded that assessment results are impacted by the stated and observed purposes of the feedback process. Ratings collected for administrative purposes were more likely to be impacted by rater biases than were ratings collected for developmental purposes. Subordinate ratings showed higher interrater reliability when the intent was positioned as being for developmental versus administrative purposes (Greguras, Robie, Schleicher, & Goff, 2003). This study showed that subordinate ratings are of significantly better quality (as defined by interrater reliability) when made for developmental than for administrative purposes, but the same is not necessarily true for peer ratings. Additional effects have also been reported, depending upon the purpose of the feedback. Bohl (1996) reports that inflated ratings are prevalent when feedback providers worry that their responses will impact a coworker’s career and pay. Farh, Cannella, and Bedeian (1991) report that peer ratings suffered from leniency and halo effects in administratively based appraisal processes.

Greguras et al. (2003) propose that there are a number of reasons why rating purpose may influence the quality of observed ratings:

1. Rating purpose may impact a rater’s motivation (e.g., political agendas and personal beliefs). For example, raters may intentionally or unintentionally bias administrative ratings to: (a) avoid the potential consequences of negative
evaluations, (b) receive the most desirable outcomes for themselves or others, or (c) motivate poor performers.

2. Different rating purposes may conflict with the rater’s role in the organization. Some raters may feel uncomfortable influencing administrative decisions that are typically made by supervisors.

3. Different raters may use different performance standards. Raters who provide administrative ratings will likely make evaluations of performance relative to other ratees. Raters who provide developmental ratings will likely make evaluations about absolute level of performance.

Opinions found in the literature regarding the use of multi-rater assessment are varied. The vast majority of published material clearly supports the use of multi-rater feedback for development. Ongoing debate exists, however, over whether it should also be used for evaluative purposes. Despite the differences of opinion found throughout various writings, it appears that many agree on the notion that most top-down, single-rater performance evaluations are inadequate in assessing true overall performance and behavior.

*Multi-rater Assessment and Performance*

It is clear that multi-rater assessment is an important tool in measuring behavioral aspects of leadership. However, there is some disagreement about how multi-rater feedback for development relates to performance on the job.
Some critics claim that some of the elements assessed in developmental multi-rater feedback are intrinsic characteristics, and will not be improved through the gathering of data. According to Van Velsor et al. (1998), traits like IQ and certain personality characteristics are innate, remaining constant over time. However, some leadership characteristics can be learned. These include: (a) self-awareness, (b) self-confidence, (c) ability to take broad, systemic view, (d) ability to work effectively in social systems, (e) ability to think creatively, and (f) ability to learn from experience. Daniel Goleman (2000) preaches a similar philosophy and list of characteristics in claiming, “Many managers mistakenly assume that leadership style is a function of personality rather than strategic choice.” He goes on to outline the following (learnable) strengths inherent in successful leaders, which include: self-awareness, self-regulation, motivation, empathy, and social skills (p. 21). These are precisely the type of factors that are assessed through developmental feedback, that are generally not included in a traditional performance evaluation.

Waldman et al. (1998) claim that many organizations have adopted multi-rater feedback in response to environmental pressures, as organizations attempt to imitate their competition or other firms. This gives a sense of external legitimacy to their process. The authors go on to cite similar examples, such as TQM and Six Sigma, and the manner in which they have been adopted. They claim that many late adopters of multi-rater feedback may not see the need to demonstrate performance improvements systematically; these assessments have become commonplace in other organizations. Therefore, these organizations do not attempt to link multi-rater feedback to concrete individual and organization performance improvement.
Others question the parallels between multi-rater feedback and performance. Smither et al. (2005), in their recent research, argue that “practitioners should not expect large, widespread performance improvement after employees receive multisource feedback… we present [that] some feedback recipients should be more likely to improve than others” (p. 33). Locke and Latham (1990) assert that feedback alone is not the cause of behavior change; it is the goals that people set in response to feedback that are the cause for change. However, there are studies that show multi-rater feedback ratings are related to performance measures (Smither et al., 2005). Several studies identify positive relationships between multisource ratings and assessment center performance (Atkins & Wood, 1995). Conway, Lombardo, and Sanders (2001), as well as Warech, Smither, Reilly, Millsap, & Reilly (1998), showed that subordinate and peer ratings were related to objective performance measures, such as production. Studies by Church (2000) provided additional insight by identifying that managers who received more favorable multisource feedback had lower turnover and higher service quality within their work teams. Smither and Walker (2001) found that feedback ratings of managers correlated with measures of customer loyalty. At the very least, multi-rater feedback may enhance two-way communication, improve relationships, uncover and resolve conflict, and provide good ideas for improvement (London & Beatty, 1993).

Waldman et al. (1998) claim that by “increasing managerial self-awareness through formalized 360 degree or upward feedback, an organization’s culture will become more participatory and will be able to react more quickly to the needs of internal and external customers.” Yet, according to these authors, “little research evidence exists...
regarding the precise methods and contexts in which it can positively affect organizational outcomes.” (p. 87)

Goleman (2000), as well as numerous others (Collins, 2001; Deming, 1992; Hannah, 1988; Jaques & Clement, 1991; Koestenbaum, 2002; Northouse, 2001) argue that effective leadership drives organization success. He states that organization climate can greatly influence financial results; as stated earlier, he believes it can account for nearly a third of financial performance. Organization climate is influenced by leadership style. As detailed in this study, developmental multi-rater feedback is intended to measure leadership behaviors and styles most critical to the success of the organization. The purpose of traditional performance appraisals, on the other hand, is to measure how successful an employee is at hitting stated performance targets; it is not designed to measure leadership style. However, each (leadership style and the ability to meet performance objectives) is important in overall success. Each, by design, measures different elements.

Although many organizations have attempted to combine the two types of feedback (traditional performance appraisal and developmental multi-rater feedback), each serves a distinct purpose, and each is important to the overall organization performance. In most organizations that have adopted multi-rater performance appraisal, the feedback is used as a supplement to supervisors' appraisals, not a replacement for them (Edwards & Ewen, 1996). As developmental multi-rater feedback and single-rater performance evaluations truly get at different outcomes, they should not be considered interchangeable elements of performance development plans. Fleenor and Prince (1997) articulate that it may be necessary to develop separate assessments for each purpose,
because feedback gathered for development is not necessarily applicable to pay or promotion decisions. Hirsch (1994) asserts that multi-rater feedback should be used for management development and coaching initiatives, while other techniques should be used for performance appraisal.

Summary

This chapter presented a review of the literature on feedback and performance management in organizations, as well as the differences between and purposes of traditional performance appraisal and developmental multi-rater assessments. This review also highlighted the expansion and growth of multi-rater assessments for both administrative (appraisal) and developmental purposes.

There are two formal feedback processes many organizations are using today: traditional top-down, single-rater performance appraisals, and developmental multi-rater assessments. Traditional performance appraisals generally serve as measurements of what targeted results were achieved. Developmental multi-rater appraisal indicates how these results were achieved. Performance appraisal is generally conducted for administrative purposes, such as pay and promotion. The intent of developmental feedback is to serve as input for an employee’s growth and development.

Most multi-rater feedback is used for developmental purposes. However, there is a growing trend to use multi-rater feedback for appraisal as well. While there is a significant amount of research on the pros and cons of using multi-rater feedback for appraisal and/or development, there is little research that addresses the correlation
between traditional performance appraisal results and developmental multi-rater feedback results for a group of employees. This study addresses these issues.
Chapter Three: Methodology

The purpose of this study was to identify the relationship developmental multi-rater feedback assessment has to top-down, traditional single-rater performance appraisals. It examines whether results of multi-rater assessments parallel those of performance appraisals.

Research Design

This study was primarily a correlation study. As defined by Kumar (1996), the intent of this type of research is to "discover or establish the existence of a relationship/association/interdependence between two or more aspects of a situation" (p. 9). The study was primarily quantitative in nature.

Research Questions

The following research questions were addressed:

1. What relationship exists, if any, between the feedback obtained via multi-rater assessments and single-rater, top-down performance appraisals?

2. Do those receiving high top-down, single-rater performance evaluation scores also receive high multi-rater scores (feedback results)? Conversely, do those receiving low single-rater performance evaluation scores also receive low multi-rater scores (feedback results)?
Research Hypotheses

This research addressed the following hypotheses:

1. **H₀₁**: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score is not statistically significant.
   
   **H₁**: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score is statistically significant.

2. **H₀₂**: Individuals with the highest multi-rater feedback scores do not receive the highest performance appraisal scores.
   
   **H₂**: Individuals with the highest multi-rater feedback scores also receive the highest performance appraisal scores.

3. **H₀₃**: Individuals with the lowest multi-rater feedback scores do not receive the lowest performance appraisal scores.
   
   **H₃**: Individuals with the lowest multi-rater feedback scores also receive the lowest performance appraisal scores.

Research Background

The organizations involved in this research are DecisionWise, Inc. and the Albuquerque, New Mexico facility of General Mills. DecisionWise is a privately held
consulting firm headquartered in Provo, Utah, providing services in organization and management development. DecisionWise’s business is based on survey administration: primarily organization assessments and multi-rater assessments. The company has been administering multi-rater assessments for approximately eight years. The majority of DecisionWise clients are Fortune 500 companies.

The Albuquerque, New Mexico facility of General Mills is a cereal manufacturing facility. This facility began construction in 1990, and became operational in the summer of 1992. At the time these assessments were completed, the facility employed 137 Cereal Manufacturing Technicians (CMTs), who serve as the focus of this study. These CMTs were supervised by eight team leaders. CMTs regularly interact with fellow CMTs, team leaders, and support personal.

General Mills began contracting with DecisionWise in 2001 to provide multi-rater assessment services to General Mills CMTs. The 2005 iteration of these multi-rater assessments marks the fifth consecutive annual administration of these surveys.

In addition to the multi-rater assessments administered by DecisionWise, General Mills was also conducting yearly performance appraisals. These performance appraisals were discontinued after the 2002-2003 administration, as the company felt the use of both performance appraisals and multi-rater assessments resulted in redundancy.

General Mills used an in-house process for performance evaluations that was designed jointly by an internal Organization Development Team and DecisionWise. The survey questions for the in-house performance appraisals were similar to those used for the multi-rater assessments; although they contained some differences. The multi-rater assessment was designed and positioned within General Mills as developmental in nature.
and purpose. The questions on this assessment were slightly different from those on the performance appraisal, in that they were intended to get at the “how” things were done (as described previously) versus the “what” was done, as in the case of the performance evaluation. Each question on the performance appraisal had a corresponding question in the multi-rater assessment.

As it does with all clients, DecisionWise has stored the data from past administrations of the General Mills multi-rater assessments. These data are owned by DecisionWise and General Mills, as per the standard Statement of Work agreed upon by both parties. General Mills has granted permission for DecisionWise to use the performance appraisal scores (written comments omitted) for research purposes, as well as to provide General Mills with the findings of this research. Existing data were taken from two assessments completed by CMTs—a multi-rater assessment and a performance evaluation—during the 2002-2003 time period. After removing CMTs that did not have both a multi-rater assessment score and a performance evaluation score, the total study population for this research was 105 participants (N=105).

Assessment Instruments

The DecisionWise multi-rater assessment is based on a survey that has been conducted by DecisionWise for over eight years with hundreds of organizations, titled the Leadership Intelligence™ 360 for Business Leaders. It consists of 37 questions, 34 of which are based on a five-point (plus “don’t know”) summated (Likert) agreement scale. This five-point scale also allows for half-point increments (1.5, 2.5, etc.) in order to provide a greater freedom of scoring range. Two questions are open-ended. The first
question on the assessment is a demographic question, used to identify whether the respondent is "self," "supervisor," "peer," or "other." Although the multi-rater survey is based on standard assessment instruments used by DecisionWise, approximately 30% of the survey content was customized specifically to measure competencies previously defined by General Mills. This multi-rater assessment instrument was designed for use as a developmental tool, rather than as an administrative or appraisal tool (see Appendix A). CMTs and their raters were notified (and had come to expect) that the multi-rater assessment was to be used solely for developmental purposes.

The multi-rater assessment, as previously mentioned, was based on a 5-point (with half-point increments) Likert scale. The following rating scales and scoring guidelines were provided to respondents:

1.0 - Poor: Consistently fails to meet requirements

1.5 - Low: Most often fails to meet requirements

2.0 – Mixed: meets some requirements, but not consistently

2.5 – Fair: Meets most requirements, but not consistently

3.0 – Solid: Regularly meets requirements

3.5 – Strong: Consistently meets and occasionally exceeds requirements

4.0 – Excellent: Regularly meets and occasionally exceeds requirements

4.5 – Exceptional: Consistently exceeds requirements

5.0 – Outstanding: The absolute role model, always exceeds requirements

0 – DK / NA: Don’t know, Not Applicable
This assessment contained nine summary categories: System Management, Work Effort/Initiative, Product Quality, Teamwork, Sanitation, Interpersonal Skills, Problem Solving, Leadership, and Safety. Each of these summary categories provided a summary mean score of all questions that fell under that particular category. Two to four separate questions were contained within each summary category. The assessment also contained three general questions about achievement, relationship, and skills, as well as open-ended comments. These general questions and open-ended comments were not used for purposes of this study.

The General Mills performance appraisal (see Appendix B) was in use for approximately 6 years, from 1997 to 2003. Its primary design purpose was administrative, as pay decisions and promotions were based largely on the results of the appraisal. The appraisal consisted of 10 questions, each rated on a scale from one to five. Each question contained an overall category descriptor, and provided descriptive sentences indicating the behaviors and accomplishments required to score high on that particular question. The category descriptors, referred to as “Performance Factors” on the General Mills appraisal, are identical in name to the summary categories on the multi-rater assessment.

Although they address the same areas in general, questions on the performance appraisal tended to focus on what was accomplished, rather than how it was accomplished, as in the case of the multi-rater assessment. For example, in the category of “Safety,” the multi-rater assessment addresses the statement, “Motivates others to behave safely by his/her example.” This question addresses the person’s leadership style and how he or she was able to motivate others in terms of safety. The corresponding
performance appraisal, on the other hand, states, "Wears bump cap, safety shoes, and hearing protection as appropriate. Does not take risks with his/her team members’ personal safety." The appraisal is more focused on what, specifically, is done. The performance appraisal also contained one more category than the multi-rater assessment. This category, labeled "Attendance," was excluded from this study.

The performance appraisal was administered by the CMT’s supervisor and was submitted to the Human Resources Department for administration of corresponding wage increases. Performance improvement plans were put in place for those employees receiving evaluation scores falling below acceptable ranges.

The multi-rater assessment grouped multiple questions into categories called "summary categories." Each summary category score was made up of the mean score of all questions (two to four) that fell under that category. Performance appraisal scores, on the other hand, represented actual scores rather than category means, as supervisors provided only one score per performance factor category.

As previously addressed in this manuscript, issues of validity and reliability may be a concern with any instrument. However, as the purpose of this research is not validation of an instrument or instruments as indicators of actual performance or behavior, issues of validity and reliability of the instruments themselves are not addressed in this study. It is important to note, however, that these instruments were designed in cooperation with a professional assessment firm (DecisionWise), have been in use for a number of years, have been used repeatedly within the manufacturing plant, and have been accepted as useful and accurate by General Mills. Because of these factors, it is assumed that these instruments are valid for purposes of this study.
Assessment Administration

Each CMT had 5-18 raters providing feedback for the Leadership Intelligence™ 360. In general, raters included Self, Supervisor, Peers, and Others. As this facility functions as an accumulation of self-directed work teams at the floor level, CMTs did not have direct reports, and therefore have no input from subordinates. This factor will be addressed later on in this study.

Each participant and corresponding rater completed the survey online through a passworded DecisionWise link. It was estimated that each survey took between 10 and 15 minutes to complete. All raters, including the CMT, were given 2 weeks to complete the survey. A reminder was sent out by DecisionWise at the 1-week point. Raters were not required to participate. However, the past four survey administrations had averaged a 90% participation rate.

Both CMTs and corresponding raters were informed through multiple methods that the developmental survey was completely confidential, was administered by a third party, and was to be used only for developmental purposes. Raters were selected by the manufacturing plant’s Human Resources Department, based on the area in which the CMT works. CMTs had the option to add raters if they chose to do so. No administrative actions (pay changes, promotion, disciplinary action, etc.) were based on the results of these assessments. Strict past adherence to this policy has proven the company’s intent and credibility in this area.

Although the multi-rater feedback was completely confidential, each employee was required to share a development action plan with his or her supervisor after
participating in the multi-rater assessment process. However, the CMT was not required to share specific feedback results with others, including his or her supervisor.

Performance appraisals were conducted annually by the CMT’s supervisor. All CMTs were required to participate in the performance review process on an annual basis. Performance appraisals were completed on paper. Overall performance ratings and performance factor ratings were then entered and tracked via MS Excel by General Mills and DecisionWise.

A total of 77% of CMTs completed both a multi-rater assessment and a performance appraisal. Of the total population of 137 CMTs, 132 completed a multi-rater assessment (96% participation rate) and 110 (82% participation rate) completed a performance review. Of these, 105 CMTs had data from both a multi-rater assessment and a performance review. This research is based on the data taken from these 105 CMTs, as they provided sufficient information for correlation at an individual level. This data represents assessments completed over the 2002-2003 time period.

Data Collection Procedures

The data was collected as follows:

1. A Statement of Work to gather both multi-rater assessment data and performance appraisal data was signed by DecisionWise and General Mills in November of 2002.

2. CMTs and their raters completed multi-rater assessments over a six-week time period. Multi-rater assessments were completed online via a secure DecisionWise website, using a standard browser and internet access.
3. Each CMT received a sealed report from DecisionWise that provided the results of their individual multi-rater assessment.

4. Supervisors completed performance appraisals on paper for each of their direct reports. This process took place over a period of approximately six months. Once complete, supervisors met with their direct reports and reviewed these evaluations.

5. A number (code) from 1-137 was assigned to each employee in order to protect identity.

6. Performance appraisal scores were matched with corresponding multi-rater assessment scores and entered into an MS Excel spreadsheet by employee code.

7. Data was made available for research purposes in 2004.

Human Subjects Considerations

This research utilizes data obtained from 2002-2004 by DecisionWise, under client contract. The actual research itself involves the analysis of pre-existing data. Data was gathered for purposes separate from this research study. This data was part of standard data gathered by both the manufacturing organization and DecisionWise. The gathering of this information is routinely considered part of the manufacturing organization’s annual, standard business practice.

Before this data was analyzed, all information which would make a participant identifiable, either as someone whose performance was being rated or who was rating someone’s performance, was removed. Numeric codes were assigned to each CMT. No
information identifying specific participants is identifiable to the researcher, DecisionWise, or client organization in this instance.

All multi-rater assessment data was gathered online via secured server. There was no interaction, other than online internet access, between the DecisionWise database and General Mills information systems. As previously indicated, multi-rater assessment results were provided directly to the CMT in a sealed cover. No individual CMT results were shared with others within General Mills, unless the CMT chose to do so voluntarily. No demographic data, other than defining the relationship raters had to the CMT, was gathered.

Performance appraisals are a part of widely accepted business practice. However, in order to mitigate risk, all employee results were coded in order to protect confidentiality. The results of this analysis will not result in any future action that will directly impact the participants of this study as a result of being subjects in the study.

This study met the criteria for exemption from full review by Pepperdine University's Graduate and Professional Schools Institutional Review Board (GPS IRB). Clearance from Pepperdine’s GPS IRB was obtained (see Appendix C).

Statistical Treatment of the Data

A number of statistical procedures were used in this analysis:

1. Measures of central tendency were used to provide descriptive analyses of multi-rater assessment and performance appraisal scores.

2. Variability and variation were calculated using standard deviation and ranges of scores.
3. A Wilcoxon signed rank test was used to test the median difference in the
   paired data.

4. Individual multi-rater overall mean scores, sorted by employee number, were
   plotted against the corresponding CMT's individual performance evaluation
   overall mean score. A similar graph was plotted by individual multi-rater
   overall scores, in descending order according to multi-rater scores, against the
   corresponding individual CMT’s performance evaluation overall score.

5. Bivariate correlation analysis was used to determine the relationship between
   overall multi-rater assessments and performance appraisal scores.
   Additionally, correlation analysis was used to identify the relationship
   between each of the multi-rater assessments’ nine summary categories and the
   corresponding nine performance factors on the performance appraisal. From
   these, Pearson’s product-moment correlation coefficients (r) of individual
   CMT multi-rater and performance evaluation scores were determined in order
   to indicate the strength and direction of the relationship between these
   variables (Holcomb, 2002). A coefficient of determination (r²) was also
   determined for each of these relationships.

6. The top 10% of all multi-rater assessment scores were analyzed against the
   top 10% of performance evaluation scores using the correlation analyses
   described above. Similar analysis was conducted for the bottom 10% of
   multi-rater and performance appraisal scores. Spearman rank-order
   correlation was used to further determine relationship of all items.
Spearman's rho ($r_s$) values were calculated in order to determine the extent of a possible relationship between these ranked scores (McCall, 2002).

7. The top 10% of multi-rater assessments were plotted (using a bar chart) against the top 10% of all performance evaluation scores to determine whether or not similarities existed in the way these CMTs scored in relationship to category summaries.

8. All statistical tests were examined for significance at the .05 level.

Analysis of data was completed using two software applications: MS Excel and SPSS.

Summary

The purpose of this study is to determine the relationship of multi-rater feedback to traditional performance evaluations. The initial research population is a group of 105 Cereal Manufacturing Technicians (CMTs) from the Albuquerque, NM facility of General Mills.

Existing data was taken from assessments administered during the 2002-2003 time period. This data was analyzed using measures of central tendency, measures of variation (standard deviations and ranges), Wilcoxon signed rank test, plotting of scores, bivariate correlation, and Spearman rank-order correlation.
Chapter Four: Results

The purpose of this study is to examine the relationship of multi-rater feedback to traditional performance evaluations. The study looked at 105 Cereal Manufacturing Technicians (CMTs) from an Albuquerque, NM food processing facility of General Mills. The data for this study was taken from existing data gathered from years 2002-2003, representing multi-rater feedback and performance appraisals for these CMTs. Data was analyzed to determine the relationship between these two forms of feedback.

This chapter will report the results of this study. It reviews research questions and hypotheses, discusses instrumentation used, details data collection methods, and provides a summary of the findings.

Research Questions and Hypotheses

The following research questions were examined:

1. What relationship exists, if any, between the feedback obtained via multi-rater assessments and single-rater, top-down performance appraisals?

2. Do those receiving high top-down, single-rater performance evaluation scores also receive high multi-rater scores (feedback results)? Conversely, do those receiving low single-rater performance evaluation scores also receive low multi-rater scores (feedback results)?

This research addressed the following null and alternate hypotheses:
1. **H01**: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score *is not* statistically significant.

   **H1**: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score *is* statistically significant.

2. **H02**: Individuals with the highest multi-rater feedback scores *do not* receive the highest performance appraisal scores.

   **H2**: Individuals with the highest multi-rater feedback scores *also* receive the highest performance appraisal scores.

3. **H03**: Individuals with the lowest multi-rater feedback scores *do not* receive the lowest performance appraisal scores.

   **H3**: Individuals with the lowest multi-rater feedback scores *also* receive the lowest performance appraisal scores.

*Assessment Instruments*

This study used data obtained from two sources, both developed together by DecisionWise and General Mills. The DecisionWise multi-rater assessment consists of 37 questions, 34 of which are based on a 5-point (plus “don’t know”) Likert scale. This
5-point scale also allows for half-point increments (1.5, 2.5, etc.) in order to provide a greater freedom of scoring range. Two questions are open-ended, and were not included in this study. The first question on the assessment is a demographic question, used to identify whether the respondent is “Self,” “Supervisor,” “Peer,” or “Other.” The multi-rater survey is based on standard assessments used by DecisionWise. However, approximately 30% of the survey content was customized specifically to measure competencies previously defined by General Mills (see Appendix A).

This multi-rater assessment contained nine summary categories. Each of these summary categories provided a summary mean score of all questions that fell under that particular category. Two to four separate questions were contained within each summary category.

The General Mills performance appraisal consisted of 10 questions, each rated on a scale from 1 to 5 (see Appendix B). Each question contained a single overall category descriptor, and provided multiple descriptive sentences indicating the behaviors and accomplishments required to score high on that particular question. One score is given for each category descriptor. The category descriptors, referred to as “Performance Factors” on the General Mills appraisal, are identical to the summary categories on the multi-rater assessment, with one exception. The performance appraisal contained an “Attendance” category descriptor. Data within this category was not used for purposes of this study, as there was no direct corresponding summary category in the multi-rater assessment with which to compare responses. The appraisal was administered by the CMT’s supervisor.
Analysis Overview

The following sections present an analysis of the data, as well as findings from this analysis. Each research question will be presented, along with corresponding null and alternate hypotheses. The analysis methodology is outlined, and results of the analyses are detailed.

Research Question 1

The following research question was examined:

What relationship exists, if any, between the feedback obtained via multi-rater assessments and single-rater, top-down performance appraisals?

There were five sub-questions to be answered that would provide answers to the first research question. These questions were:

1.1 Are multi-rater overall mean, median, and mode scores different from the performance evaluations’ mean, median, and mode scores?

1.2 Is there a significant difference in the standard deviations between multi-rater scores and performance evaluation scores?

1.3 Is the range of scores within the multi-rater feedback different from the performance evaluation overall scoring range?

1.4 Do Pearson product-moment correlation coefficients \(r\) indicate a significant correlation between overall multi-rater feedback scores and overall performance evaluation scores?
1.5 Do Pearson product-moment correlation coefficients (r) indicate a significant correlation between the nine summary categories within the multi-rater assessments and the nine performance factors within the performance appraisals?

Based on the number of participants (N=105), a difference of 15% in rater scores in the above sub-questions is considered significant for purposes of this study. Significant differences found in three or more of the above sub-questions affirm a significant difference in answering Research Question 1 overall.

The following hypotheses were tested in association with Research Question 1:

Ho1: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score is not statistically significant.

H1: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score is statistically significant.

The first sub-question (1.1) to be answered under this category was whether multi-rater overall mean, median, and mode scores differed from the performance evaluation mean, median, and mode scores. Measures of central tendency were used to provide descriptive analyses of multi-rater assessment and performance appraisal scores.

The mean for the nine multi-rater feedback category summary scores was found to be 4.4, while the mean for the nine performance appraisals performance factors
was 3.6. This indicated a gap in assessments of 0.8 on a five-point scale when comparing mean to mean. This equates to a significant difference of 16%.

The median for each assessment was then studied. The median score for multi-rater feedback, using the same categories as before, was 4.5. A median performance evaluation score was found to be 3.6—a significant difference of 0.9 over a five-point scale—which equates to 18%.

Next, the modes for multi-rater assessments and performance evaluations were evaluated. Calculations showed a mode for multi-rater assessments of 4.6. The mode for performance appraisals was 3.0. A difference of 1.6 over a five-point scale was found, resulting in a gap of 32%. While this difference is significant, it may be somewhat distorted in that the multi-rater score is based on a mean score of all questions that fall within a multi-rater summary category. In other words, the summary category score is the mean of two to four question scores. As a result, the score may be expressed in fractions of a whole number (e.g., 4.6). Performance evaluation scores, on the other hand, are mean scores of nine performance factors for all CMTs. These performance factors do not represent mean scores of sub-questions falling under the performance factor, and are therefore expressed only in whole numbers. This does not allow for scores between whole numbers on individual performance factors. Thus, performance factor category scores on performance evaluations may be less concise than multi-rater summary category scores.

These scores and gaps are expressed graphically in Figure 1. These findings show a clear difference in measures of central tendency. In each case, multi-rater feedback measures were significantly higher than performance evaluation descriptors, showing
a 16% difference in the mean scores, an 18% difference in medians, and a mode that was 32% higher in multi-rater scores than it was in performance evaluation scores.

Figure 1. Mean, median, and mode for performance appraisal and multi-rater assessments.

The next sub-question (1.2) to be answered for Research Question 1 was if there was a statistically significant difference (15% or greater) between the standard deviations of multi-rater feedback and performance evaluation scores. Analysis again showed a significant difference, as the standard deviation for multi-rater assessment scores was 0.30 and the standard deviation for performance evaluations was 0.46 (see Table 1).
Table 1
Mean, Median, Mode, and Standard Deviation

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-rater assessments</td>
<td>4.4</td>
<td>4.5</td>
<td>4.6</td>
<td>0.30</td>
</tr>
<tr>
<td>Performance Evaluations</td>
<td>3.6</td>
<td>3.6</td>
<td>3.0</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Ranges (sub-question 1.3) were then calculated between overall multi-rater feedback scores and performance evaluation scores. The highest individual multi-rater score was 4.88, with a low score of 3.33. This difference represents a range of 1.55. Performance evaluation scores ranged from a high of 4.67 to a low of 2.66, indicating a range of 2.01. This difference between assessment ranges (1.55 and 2.01) represents a difference in range of 0.46, or 9% over a five-point scale. Although is a clear difference in ranges (see Table 2), this difference was not found to be significant (15% or greater).

Table 2
Ranges of Multi-rater Overall Scores vs. Performance Evaluation Overall Scores

<table>
<thead>
<tr>
<th></th>
<th>Multi-rater Overall Score</th>
<th>Performance Evaluation Overall Score</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest individual score</td>
<td>4.88</td>
<td>4.67</td>
<td>0.21</td>
</tr>
<tr>
<td>Lowest individual score</td>
<td>3.33</td>
<td>2.66</td>
<td>0.67</td>
</tr>
<tr>
<td>Range</td>
<td>1.55</td>
<td>2.01</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
A Wilcoxon signed rank test was used to test the median difference in the paired data. This test, with the ranking of overall performance evaluation scores against overall multi-rater assessment scores, showed 99 negative ranks and 6 positive ranks, indicating that only six CMTs received higher scores on their performance reviews than on their multi-rater feedback. Conversely, 99 CMTs received higher scores on their multi-rater assessments than they received on their performance appraisals. The Wilcoxon Z statistic, which is a standardized measure of the distance between the rank sum of the negative group and its expected value (the negative group rank sum, minus half the sum of all ranks, divided by the standard deviation) is -8.755. This test shows that in all but six of the 105 cases, multi-rater feedback scores were higher than performance rating scores. Six percent of the CMTs received higher performance evaluation scores than multi-rater assessment scores, while 94% of CMTs received higher multi-rater scores than performance evaluation scores—a significant difference greater than 15%.

The next sub-question (1.4) that was posed in order to address Research Question 1 was what specific correlation exists between overall CMT multi-rater scores and CMT performance evaluation scores? Individual scores were sorted and ordered by coded employee number from 1 to 105. Corresponding multi-rater and performance appraisal results (mean scores for each participant) were then plotted (see Figure 2). Visual inspection of the results indicate that overall mean multi-rater scores exceed overall mean performance evaluation scores in nearly all cases. Any further relationship between scores, however, could not be determined based on visual review of the plotted scores.
Figure 2. Multi-rater feedback scores plotted against performance evaluation scores (sorted by coded employee number).

Figure 3. Multi-rater feedback scores plotted against performance evaluation scores (sorted by highest to lowest multi-rater scores).
From visual inspection of Figure 3, it is clear that there is a relationship when data were plotted with the high to low multi-rater mean scores versus the performance evaluation scores. In all but six instances, multi-rater feedback scores exceeded that of the individual’s performance evaluation score. In one case they were equal in value. However, as with Figure 2, no further relationship between scores could be determined based on the examination of the plotted scores.

Correlation analyses (sub-question 1.4) were performed in order to further determine the relationship between multi-rater assessments and performance evaluations. Additional correlation analyses (sub-question 1.5) were performed to examine the relationship each of the nine categories contained within the multi-rater assessments had to the corresponding nine performance factors within the performance appraisal. Results of these analyses are show in Tables 4 and 5. Correlation coefficient values from the analysis were interpreted according to the values defined in Table 3 (Hinkle, Weirsman, & Jurs, 1979; Holcomb, 2002).

Table 3

<table>
<thead>
<tr>
<th>Value of $r$</th>
<th>Degree of Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90 to 1.00 (-0.90 to -1.00)</td>
<td>Very high positive (negative)</td>
</tr>
<tr>
<td>0.70 to 0.90 (-0.70 to -0.90)</td>
<td>High positive (negative)</td>
</tr>
<tr>
<td>0.50 to 0.70 (-0.50 to -0.70)</td>
<td>Moderate positive (negative)</td>
</tr>
<tr>
<td>0.30 to 0.50 (-0.30 to -0.50)</td>
<td>Low positive (negative)</td>
</tr>
<tr>
<td>0.00 to 0.30 (0.00 to -0.30)</td>
<td>Little, if any, correlation</td>
</tr>
</tbody>
</table>
Correlation coefficients were calculated for overall mean multi-rater scores and overall mean performance evaluation scores (sub-question 1.4). A Pearson's product-moment correlation coefficient ($r$) was determined, with an $r$ value of 0.22. This $r$ value and coefficient of determination of $r^2=0.05$ indicate that there is little, if any, correlation between individuals' overall mean multi-rater scores and overall mean performance evaluation scores for this group of 105 CMTs.

Table 4
Correlation Coefficients ($r$) and Coefficients of Determination between Multi-rater and Performance Appraisal Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>$r$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Management</td>
<td>0.30</td>
<td>0.09</td>
</tr>
<tr>
<td>Work Effectiveness/Initiative</td>
<td>0.28</td>
<td>0.08</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>0.28</td>
<td>0.08</td>
</tr>
<tr>
<td>Product Quality</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>Teamwork</td>
<td>0.27</td>
<td>0.07</td>
</tr>
<tr>
<td>Safety</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Sanitation</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>Interpersonal/Coaching Skills</td>
<td>0.35</td>
<td>0.12</td>
</tr>
<tr>
<td>Leadership</td>
<td>0.19</td>
<td>0.04</td>
</tr>
<tr>
<td>Overall</td>
<td>0.22</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note: In no instance was correlation found to be significant at the 0.05 level (two-tailed).
Correlation analyses were also performed to examine the relationship between the nine summary categories within the multi-rater assessments and the nine performance factors within the performance appraisals (sub-question 1.5). As indicated in Table 4, little to no positive correlation was found in any of the nine categories, nor in the “Overall” category.

Significant differences were found in mean median and mode scores (sub-question 1.1), as well as in the standard deviations of these scores (sub-question 1.2). Although differences were found in ranges (sub-question 1.3), these differences were not found to be significant according to the criteria within this study. Pearson product-moment correlation coefficients (sub-question 1.4) indicated that there was little, if any, correlation in overall scores, nor was there significant correlation found in the nine category scores. Based on the significance criteria described in this research (three of five sub-questions), a significant difference was found between multi-rater assessment scores and performance evaluation scores (Research Question 1).

**Research Question 2**

A second research question was then examined:

Do those receiving high top-down, single-rater performance evaluation scores also receive high multi-rater scores (feedback results)? Conversely, do those receiving low single-rater performance evaluation scores also receive low multi-rater scores (feedback results)?
The following hypotheses were tested in association with Research Question 2:

Ho2: Individuals with the highest multi-rater feedback scores do not receive the highest performance appraisal scores.

H2: Individuals with the highest multi-rater feedback scores also receive the highest performance appraisal scores.

Ho3: Individuals with the lowest multi-rater feedback scores do not receive the lowest performance appraisal scores.

H3: Individuals with the lowest multi-rater feedback scores also receive the lowest performance appraisal scores.

In order to answer this research question, Spearman rank-order correlation was used to further determine relationship of all items. Spearman’s rho ($r_s$) values were calculated in order to determine the extent of a possible relationship between these ranked scores (McCall, 2002).

The Spearman rank-order correlation test failed to confirm a significant correlation between ranked individual multi-rater scores and performance appraisal scores. Results from this test are detailed in Table 5.
Table 5
*Spearman Rank-order Correlation: Multi-rater Assessment and Performance Appraisal*

<table>
<thead>
<tr>
<th>Category</th>
<th>Spearman’s rho ($r_s$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Management</td>
<td>0.26</td>
</tr>
<tr>
<td>Work Effectiveness/Initiative</td>
<td>0.27</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>0.31</td>
</tr>
<tr>
<td>Product Quality</td>
<td>0.10</td>
</tr>
<tr>
<td>Teamwork</td>
<td>0.28</td>
</tr>
<tr>
<td>Safety</td>
<td>0.01</td>
</tr>
<tr>
<td>Sanitation</td>
<td>0.13</td>
</tr>
<tr>
<td>Interpersonal/Coaching Skills</td>
<td>0.33</td>
</tr>
<tr>
<td>Leadership</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td><strong>0.20</strong></td>
</tr>
</tbody>
</table>

*Note: In no instance was correlation found to be significant at the 0.05 level (two-tailed).*

In order to further investigate this relationship, the top 10% and bottom 10% of individual multi-rater scores were identified and ranked from highest multi-rater score to lowest multi-rater score. They were then matched against their corresponding performance evaluation rankings. The percentiles in which the highest and lowest multi-rater scores fell within their corresponding performance evaluation rankings were then analyzed (see Table 6).
Table 6  
*Top 10% and Bottom 10% of Multi-rater Rankings Compared to the Same Individuals' Performance Evaluation Rankings*

<table>
<thead>
<tr>
<th>Multi-rater Rankings</th>
<th>Top 10% of Multi-rater Rankings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-rater Ranking</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Perform. Appraisal</td>
<td>10 4 46 31 32 2 52 5 47 55</td>
</tr>
<tr>
<td>Percentile</td>
<td>90 96 56 70 70 98 50 95 55 48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bottom 10% of Multi-rater Rankings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-rater Ranking</td>
</tr>
<tr>
<td>Perform. Appraisal</td>
</tr>
<tr>
<td>Percentile</td>
</tr>
</tbody>
</table>

Analysis of Table 6 shows that of the 10 highest individual multi-rater feedback scores, 9 fell within the top 50<sup>th</sup> percentile of performance evaluation scores. Of the 10 lowest individual multi-rater feedback scores, 8 of the 10 fell within the lowest 50<sup>th</sup> percentile of performance evaluation scores. While no correlation analysis was attempted, the data clearly show that those CMTs scoring within the top 10% of multi-rater feedback
also scored in the top half of performance evaluation scores 90% of the time. Additionally, those CMTs scoring within the bottom 10% of multi-rater feedback scored in the bottom half of the performance evaluation scores 80% of the time.

It was also found that 40% (n=4) of those CMTs in the top 10% of multi-rater scores were also found in the top 10% of performance evaluation scores. This left 60% (n=6) that were found in the top 10% of multi-rater feedback scores, but not in the top 10% of performance evaluation scores. When completing a similar analysis for the bottom 10%, it was found that only 20% (n=2) of those CMTs scoring in the bottom 10% of multi-rater scores also scored in the bottom 10% of performance evaluations scores. These results indicate that the highest and lowest performers, as ranked by the top 10% and bottom 10% of multi-rater scores, were not generally identified as the highest or lowest in corresponding performance evaluation scores.

Additional analysis was completed in order to identify whether the overall category summary scores of those CMTs scoring in the top 10% received similar ratings in category summary multi-rater scores as they received in their overall performance dimensions performance appraisal scores. It was found that, similar to what was found in the overall scores of the CMT population, differences existed between multi-rater category summary scores and performance evaluation performance dimension scores. These differences were more profound in some categories than in others. For example, a difference of 0.89 on the 5-point scale was found in the Safety summary category between the CMTs with the top 10% of multi-rater scores (score= 3.8) and their corresponding performance evaluation scores (score= 4.69). Overall Systems Management category scores, on the other hand, were identical at 4.85 for both multi-
rater and performance evaluation scores (see Figure 4). Applying the 15\% significance decision rule discussed previously, only one of these nine categories was found to have a significant difference in multi-rater feedback scores and performance evaluations scores. However, seven of the nine categories showed higher multi-rater feedback scores than performance evaluation scores, with one category showing equal scores.

Figure 4. Category summary mean scores for CMTs ranking in the top 10\% of multi-rater scores.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
A similar analysis was completed on the bottom 10% of CMT scores and their corresponding performance evaluation summary scores (see Figure 5). These scores were identified and analyzed by category. Again, significant differences were found between multi-rater scores and performance evaluation scores. Once again applying the 15% significance decision rule discussed previously, five of the nine categories showed a significant difference between multi-rater feedback scores and performance evaluations scores. All nine categories also showed higher multi-rater feedback scores than performance evaluation scores.

Figure 5. Category summary mean scores for CMTs ranking in the bottom 10% of multi-rater scores.
These findings suggest that multi-rater assessments tend to be more closely related to performance evaluations when rating top performers than when rating the lowest performance performers.

**Summary of Results**

This chapter analyzed the results of two research questions. The results are summarized below:

*Research Question 1:* What relationship exists, if any, between the feedback obtained via multi-rater assessments and single-rater, top-down performance appraisals?

Results from these analyses led to accepting the following null hypothesis (H01) and thus rejecting the alternate hypothesis (H1):

H01: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score is not statistically significant.

H1: Correlation between the individual multi-rater feedback score and the individual annual performance appraisal score is statistically significant.

These analyses found the following, which led to the above conclusion:

1. There were significant differences between the mean, median, mode, and standard deviations of multi-rater feedback scores when compared to
performance evaluation scores. In each case, the measures of central tendency found on the individual multi-rater assessments were significantly higher than those found on performance evaluations.

2. Differences in ranges between individual multi-rater assessment mean scores and performance appraisal overall mean scores were significant.

3. A Wilcoxon signed rank test clearly showed that individual multi-rater feedback scores were in most cases higher than corresponding performance appraisal scores.

4. Two-tailed Pearson’s product-moment correlation tests show that little, if any correlation exists between overall multi-rater scores and overall performance appraisal scores.

5. Further correlation analysis between the nine categories on multi-rater assessments and performance evaluations failed to show significant correlation. Only one category, Interpersonal/Coaching Skills, showed a slight correlation.

Research Question 2: Do those receiving high top-down, single-rater performance evaluation scores also receive high multi-rater scores (feedback results)? Conversely, do those receiving low single-rater performance evaluation scores also receive low multi-rater scores (feedback results)?

Results from these analyses led to accepting the following second null hypothesis (Ho2), and thus rejecting the second alternate hypothesis (H2):
Ho2: Individuals with the highest multi-rater feedback scores do not receive the highest performance appraisal scores.

H2: Individuals with the highest multi-rater feedback scores also receive the highest performance appraisal scores.

These analyses also lead to accepting the following third null hypothesis (Ho3), and rejecting the third alternate hypothesis (H3):

Ho3: Individuals with the lowest multi-rater feedback scores do not receive the lowest performance appraisal scores.

H3: Individuals with the lowest multi-rater feedback scores also receive the lowest performance appraisal scores.

These analyses found the following, which led to the above conclusions:

1. The Spearman rank-order correlation test failed to confirm a correlation between individual ranked multi-rater scores and performance appraisal scores.

2. Only 40% of those CMTs whose scores were in the top 10% of multi-rater scores were also in the top 10% of performance evaluation scores.

3. Only 20% of those CMTs whose scores were in the bottom 10% of multi-rater scores were also in the bottom 10% of performance evaluation scores.
4. Those CMTs scoring within the top 10% of multi-rater feedback scores generally scored within the top half of performance evaluation scores.

5. Those CMTs scoring within the bottom 10% of multi-rater feedback scores generally scored within the bottom half of performance evaluation scores.

6. CMT scores for top performers showed a greater relationship in some summary categories than in others. In six of the nine categories, multi-rater scores were within 0.5 points (over a five-point scale) of performance evaluation scores for those scoring within the top 10% of multi-rater scores.

7. CMT scores for the lowest performers also showed a greater relationship in some summary categories than in others. When analyzing the lowest 10% of multi-rater scores, differences in category scores were all greater than 0.5 points (over a five-point scale) when compared to the same CMTs’ performance evaluations.

8. Multi-rater assessments scores tend to be more closely related to performance evaluations when rating top performers than when rating the lowest performers.

Based on this research, it appears that significant differences exist between multi-rater feedback and performance appraisal scores for this group of 105 CMTs. These two forms of assessment show little, if any, correlation. Additionally, top performers on individual multi-rater assessments did not generally fall within the highest ranks of performance evaluation scores, and the lowest performers in one assessment were not necessarily the lowest performers identified by the other assessment.
Chapter Five: Summary, Conclusion, and Recommendations

This chapter includes a review of the problem, purpose, and methodology used in this research study. It provides a review of the research questions, as well as the findings from this study. Implications and limitations of this study are discussed. Recommendations for future research and study are also provided.

Problem and Purpose

Multi-rater feedback is being used increasingly within organizations as a tool for employee development. There is currently a great deal of controversy, however, as to whether multi-rater feedback should also be used for performance appraisal purposes, or used exclusively for purposes of employee development. Understanding the relationship, if any, between multi-rater feedback used for development purposes and traditional performance evaluations would provide a better understanding of how these two components should be used jointly, separately, or not at all in developing employees within an organization.

The purpose of this study is to identify the relationship individual developmental multi-rater feedback assessments have to top-down, traditional single-rater performance appraisals. This study reviewed the individual multi-rater assessment scores in relationship to the performance appraisal scores of Cereal Manufacturing Technicians within the Albuquerque, New Mexico division of a food processing facility of General Mills.
Methodology

The population for this study was a group of 105 Cereal Manufacturing Technicians (CMTs) at a General Mills Albuquerque, NM manufacturing facility. During 2003-2004, these CMTs participated in a multi-rater assessment process, as well as a traditional performance appraisal. Although these two assessments contained nine similar overall categories, each contained slightly different questions. While the performance appraisal primarily focused on what was done, the multi-rater assessment was designed to focus on how it was done.

This research involves a correlation study conducted in 2005, reviewing existing 2002-2003 data from this group of CMTs. This data was analyzed using measures of central tendency, score ranges, measures of variation (standard deviations and ranges), Wilcoxon signed rank test, plotting of scores, bivariate correlation, and Spearman rank-order correlation. Results of this analysis were presented in chapter four of this document.

Discussion: Research Question 1

The first research question addressed the following:

R1: What relationship exists, if any, between the feedback obtained via multi-rater assessments and single-rater, top-down performance appraisals?

Results clearly show that mean, median, and mode scores tend to be significantly higher with multi-rater feedback than in performance appraisal scores. Results also
showed a significant difference in standard deviations between overall scores. Performance evaluations showed greater variability in scores than did multi-rater feedback scores. Although there was a greater range found in performance evaluation scores than in multi-rater feedback scores, this range was not found to be significant according to the established criteria of a 15% difference or greater.

These findings support studies that claim that employees indicate that they would have rated their boss differently had the feedback been used for the manager’s performance appraisal (Bohl, 1996; London et al., 1990). These findings also call into question the assumption that individual multi-rater assessment scores, in general, parallel performance appraisal scores.

This research also provides insight into differences in ranges between multi-rater feedback scores and performance evaluation scores. Although individual multi-rater scores are significantly higher than performance appraisal scores, the range for performance appraisal scores is wider. This difference may be partially accounted for when one considers that there may be a bias for ranking employees lower on performance evaluations than on multi-rater assessments, as supervisors are often required to implement a “forced ranking” system, rating employees on a bell-curve (Anonymous, 2002; Johnson, 2004; J. A. Segal, 1992). This forced ranking may also cause performance evaluators to spread scoring ranges out to the point that they do not indicate true performance levels. In other words, under this scenario, an employee who is performing his or her job to standard may, in fact, be rated as not meeting expectations because the supervisor is forced to artificially distribute performance scores across a broader rating scale. Additionally, many performance appraisal processes are also associated with
administrative actions (compensation, promotion, disciplinary action, etc.), and managers may be required to force an artificial distribution in performance evaluation scores in order to meet associated administrative guidelines.

The results of this research also indicate that little, if any, direct correlation exists between overall multi-rater feedback scores and performance evaluation scores. This may be a result of two possibilities: (a) Supervisors (those providing the performance evaluation scores) may have a different view of performance than others (those additional raters providing the multi-rater feedback); and/or (b) multi-rater feedback and performance appraisals have inherently distinct outcomes.

The first of these possibilities, the notion that supervisors and others may not agree on performance dimensions, is extensively supported by the literature (Campbell & Lee, 1988; Harris & Schaubroeck, 1988; Lublin, 1994; Yammarino & Atwater, 1993). Most organizations assume that supervisory ratings accurately capture true performance (Campbell & Lee, 1988). However, differences in the way performance is viewed or witnessed on a daily basis may vary according to the interaction one has with the person being rated—in this case, the CMT. It is also possible, if not probable, that a supervisor has different requirements of the employee than do others interacting with that employee.

The second possibility listed above, the theory that multi-rater feedback and performance appraisals have distinct outcomes, may have significant meaning to organizations. There have been numerous claims that multi-rater feedback should be purely developmental in nature (Harvey, 1994; London & Smither, 1995; London et al., 1990). However, many organizations are now using multi-rater assessment for performance appraisal (DeNisi & Kluger, 2000; London & Smither, 1995). The findings
of this study indicate that a comparison of individual multi-rater scores to individual performance appraisal scores shows little, if any correlation. Multi-rater assessments and performance appraisals provide different outcomes and are, therefore, not interchangeable.

**Discussion: Research Question 2**

The second research question addressed the following:

R2: Do those receiving high top-down, single-rater performance evaluation scores also receive high multi-rater scores (feedback results)? Conversely, do those receiving low single-rater performance evaluation scores also receive low multi-rater scores (feedback results)?

Results of this study show that top performers on individual multi-rater assessments did not generally fall within the highest ranks of performance evaluation scores. Similarly, the lowest performers in one assessment were not necessarily the lowest performers identified by the other assessment. However, top scorers (top 10%) on multi-rater assessments tended to score in the top half of performance evaluation scores. Likewise, those receiving the lowest multi-rater scores (bottom 10%) also tended to be rated within the bottom half of performance evaluation scores.

Multi-rater assessment scores were more closely related to performance evaluations when rating top performers than when rating the lowest performers. This may be due, in part, to the matter of scoring range discussed previously. As the range of
individual multi-rater scores was narrower and higher than performance evaluation scores, it follows that higher performance evaluation scores would relate more closely to their corresponding individual multi-rater scores than would the lowest scores. However, this relationship is worth further study.

When looking at summary categories, agreement between individual multi-rater scores and performance evaluation scores for top performers was also more significant in some categories than in others. This would appear to indicate that in some categories there is a more congruent view of what constitutes high performance than in other categories.

When these results are paired with the lack of correlation found through Research Question 1, we see that individual multi-rater assessments and performance evaluations clearly provide different perspectives on the evaluation of performance. There appears to be significant disagreement between the two assessments as to what constitutes the highest and lowest levels of performance, as well as who falls within those categories of performance.

Limitations of this Study

There are three primary limitations in this study. The first of these is the nature of the study population. Few opportunities exist to study an in-tact population that has completed both multi-rater assessments and performance evaluations over a short time-period. Even rarer is finding a group that has used similar instruments and methodologies. Although this study dealt with 105 CMTs, the study population is relatively small. All CMTs belonged to the same division, worked in the same location,
and performed the same basic job duties. While this homogeneity provided for a cleaner comparison by reducing the number of variables, it also reduces the ability to extrapolate and generalize these findings to a population outside of this team.

Second, because data were not available, scores provided by supervisors on the multi-rater assessment and those provided by the supervisor on the performance appraisal could not be compared to each other. The inability to make this comparison limited the ability to identify to what degree the variation in scores and lack of correlation could be accounted for by the ratings of peers, self, and others. This study did not show (nor did it attempt to do so) whether supervisors provided similar ratings on both assessments.

Third, because these CMTs had no direct reports, no direct report scores were included as part of this study. While each CMT received ratings from between 8 and 15 raters, this individual multi-rater assessment differed from many multi-rater assessments in that it did not included a subordinate rating.

**Implications of this Study**

The use of multi-rater feedback continues to increase. Rather than using multi-rater feedback strictly for employee development, many organizations now use multi-rater feedback for appraisal purposes. However, if these two forms of assessment yield similar results, it is in the best interest of organizations to reduce redundancy, time, and costs by using just one instrument for development and appraisal purposes. If these two instruments provide different outcomes, and both provide valuable insight into an employee’s behavior, skills, and results, both should be taken into account when looking at an employee’s overall development plan.
This study shows that there is a significant difference between results found in individual multi-rater assessments and those found in individual performance appraisals. It also shows that there is a tendency for those who are seen as either top or bottom performers, as indicated by a multi-rater assessment, to be also rated as respectively above or below average by performance appraisals. Based on this study, however, there is little further correlation that exists between individual multi-rater scores and performance evaluation scores.

Multi-rater assessments and performance evaluations work together to provide a more comprehensive picture of an employee’s performance and behaviors. These two assessments measure different elements. This research shows that a supervisor’s view of performance may be somewhat limited, in that he or she may not see all dimensions of performance and behavior that others might. This finding brings into question the common practice found in many organizations of relying exclusively upon supervisors to provide an evaluation of employee performance, as the evaluation may be incomplete without multiple sources of feedback. It also highlights the possibility that forced performance rankings—forcing the distribution of employee performance ratings across a five-point scale—may result in an inaccurate picture of true performance. Organizations making administrative decisions based solely on supervisor ratings should, according to this research, reconsider their methodology.

This study also questions the practice of replacing top-down performance appraisals with multi-rater assessments, as this research shows they are not interchangeable. Similarly, a traditional performance appraisal, when used alone, does not provide sufficient information from which to base an employee development plan. A
comprehensive employee performance and development plan should include both a multi-rater assessment, designed specifically for development, and a performance appraisal designed specifically for evaluation of on-the-job performance.

Recommendations for Further Study

While this research provided insight into the relationship between individual multi-rater assessments and performance evaluations, further study is recommended. Further study should include:

1. Conducting additional studies with a larger population. This should include multiple organizations, industries, positions, and levels. It is also recommended that multiple administrations of this process be conducted and evaluated over time.

2. Evaluating the relationship of supervisor ratings on multi-rater assessments to supervisor ratings on performance appraisals.

3. Conducting studies with a research population that includes employees with direct reports.

Summary

This study examined the relationship of individual multi-rater assessments to traditional top-down performance appraisal. The study reviewed the individual multi-rater feedback scores of 105 Cereal Manufacturing Technicians, and compared those scores to their corresponding performance evaluation scores. The study showed that multi-rater scores were significantly higher than performance evaluation scores. It was
also found that little, if any correlation exists between overall multi-rater feedback scores and performance evaluation scores. The study showed that these two forms of assessment provided different outcomes. As both provide valuable information, it is recommended that both a multi-rater (developmental) assessment and a performance (evaluative) appraisal be incorporated into an employee's overall performance and development plan.
References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix A: Multi-rater Feedback Form

Note: This multi-rater assessment was originally an HTML webpage document. As it was designed for display on a computer monitor rather than as a printed document, this form is not formatted as it would be when displayed electronically on a monitor.
General Mills Peer Review Survey

This survey provides feedback to a person on areas critical to the person's success at General Mills. The statements listed below describe the behaviors required for performance excellence. It is important that you provide thoughtful and candid feedback. Provide the complete name of the person requesting feedback. Without this information, we cannot include your responses in the feedback report. When you're done, submit your responses by clicking "Submit" below. Thanks for your participation!

You are giving feedback for: $user_name

1. You are giving feedback as . . .
   ○ Self- I am the person providing feedback on myself.
   ○ Supervisor- I am the supervisor of the person requesting feedback.
   ○ Peer- I am the peer of the person requesting feedback.
   ○ Other- I work with the person requesting feedback in some other relationship than the ones described above.

Instructions:
Indicate how effective the person requesting feedback is in each of the following competencies/areas. Note that a Solid rating (3.0) is a favorable score. A Strong rating (3.5) is a significant achievement given General Mills' high performance standard. The highest rating (5.0) is a very rare rating given only for those who always exceed requirements.

1.0 - Poor: Consistently fails to meet requirements.
1.5 - Low: Most often fails to meet requirements.
2.0 - Mixed: Meets some requirements, but not consistently.
2.5 - Fair: Meets most requirements, but not consistently.
3.0 - Solid: Regularly meets requirements.
3.5 - Strong: Consistently meets and occasionally exceeds requirements.
4.0 - Excellent: Regularly exceeds requirements.
4.5 - Exceptional: Consistently exceeds requirements
5.0 - Outstanding: The absolute role model, always exceeds requirements.
0 - DK / NA: Don't Know, Not Applicable

1 - Poor , 1.5 - Low , 2 - Mixed , 2.5 - Fair , 3 - Solid , 3.5 - Strong , 4 - Excellent , 4.5 - Exceptional , 5 - Outstanding , 0 - Don't Know / Not applicable

1 2 3 4 5 0

System Management | | | | | |
2. Operates his/her system to maximize system performance.
3. Manages his/her work with precision and an eye for detail.
4. Manages work in a way that facilitates upcoming events (like changeovers, hand-offs, etc.).
5. Quickly corrects potential and real issues on her/his system.

1 - Poor, 1.5 - Low, 2 - Mixed, 2.5 - Fair, 3 - Solid, 3.5 - Strong, 4 - Excellent, 4.5 - Exceptional, 5 - Outstanding, 0 - Don't Know / Not applicable

**Work Effort / Initiative**
6. Takes personal responsibility for his/her own performance.
7. Manages her/his time effectively with regard to breaks and arrival for hand-offs.
8. Consistently makes a positive contribution towards both daily production and longer-term plant goals.
9. Eliminates waste and improves efficiency in every aspect of work.
10. Is constantly learning the necessary skills to do the job.

1 - Poor, 1.5 - Low, 2 - Mixed, 2.5 - Fair, 3 - Solid, 3.5 - Strong, 4 - Excellent, 4.5 - Exceptional, 5 - Outstanding, 0 - Don't Know / Not applicable

**Product Quality**
11. Makes a strong effort to improve product quality.
12. Is quick to respond to and correct product quality issues.

1 - Poor, 1.5 - Low, 2 - Mixed, 2.5 - Fair, 3 - Solid, 3.5 - Strong, 4 - Excellent, 4.5 - Exceptional, 5 - Outstanding, 0 - Don't Know / Not applicable

**Teamwork**
13. Motivates team members to cooperate and help each other.
14. Communicates clearly with those whose jobs s/he directly affects.
15. Actively participates in team discussions and decision-making.
16. Takes personal accountability for the results of his/her team.

1 - Poor, 1.5 - Low, 2 - Mixed, 2.5 - Fair, 3 - Solid, 3.5 - Strong, 4 - Excellent, 4.5 - Exceptional, 5 - Outstanding, 0 - Don't Know / Not applicable

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Sanitation
17. Maintains a neat and sanitary work area.
18. Contributes to daily and downtime sanitation efforts.
19. Demonstrates a clear understanding of the importance of plant sanitation guidelines including GMP's.

Interpersonal Skills
20. Builds strong relationships with others.
21. Fosters energy, enthusiasm and commitment in others.
22. Handles stressful or changing circumstances with a positive attitude.
23. gladly accepts and incorporates feedback which results in greater performance.

Problem Solving
24. Demonstrates good judgement and common sense when facing a problem.
25. Demonstrates effective troubleshooting skills and identifies root causes.
26. Considers multiple priorities and sources of information before making important decisions.

Leadership
27. Participates in training others to improve performance.
28. Encourages others, by example, to perform at a higher level.
29. Discusses with team members how current performance contributes to larger objectives.

Safety
30. Demonstrates understanding of how job performance

1 - Poor, 1.5 - Low, 2 - Mixed, 2.5 - Fair, 3 - Solid, 3.5 - Strong, 4 - Excellent, 4.5 - Exceptional, 5 - Outstanding, 0 - Don't Know / Not applicable
contributes to team and plant goals.

31. Motivates others to behave safely by his/her example.

32. Takes appropriate action to correct unsafe working conditions.

1 - Poor, 1.5 - Low, 2 - Mixed, 2.5 - Fair, 3 - Solid, 3.5 - Strong, 4 - Excellent, 4.5 - Exceptional, 5 - Outstanding, 0 - Don't Know / Not applicable

Overall

33. Overall, I rate this person's contributions and achievements as . . .

34. Overall, I rate this person's relationships with others as . . .

35. Overall, I rate this person's skills and abilities as . . .

36. Describe this person's greatest strengths as a team member.

37. Describe specific things this person could do to become a better team member.

* We encourage you to include your name with the survey because it is more meaningful. However, it is not required.

Press the SUBMIT button. Thank you for taking the survey.
Appendix B: Performance Appraisal Form
**GENERAL MILLS, INC.**  
Technology and Operations

Performance Planning and Review  
Team Member

<table>
<thead>
<tr>
<th>Employee</th>
<th>Current Position</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team Leader</td>
<td>Team</td>
<td></td>
</tr>
</tbody>
</table>

**PERFORMANCE FACTORS**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SYSTEM MANAGEMENT – Manages and operates the system and equipment in a productive manner. Strives to operate the system at maximum operating rates that also deliver quality product. Plans ahead to ensure PM’s are done on time and equipment is set up for the next team. Plans ahead for shutdowns and cleanouts. Strives to minimize losses, overpack, certa, and offal.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rating</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. WORK EFFORT/INITIATIVE – Puts forth the effort to do the complete job. Sets up the next team with a clean work area; with all systems running and QC checks done. Can be counted on by the team to carry his/her work load. Doesn’t overly rely on team members to cover for him/her. Uses idle time effectively to do sanitation and PMs. Responds to problems in an urgent manner. Manages break time in a responsible manner.</td>
<td></td>
</tr>
</tbody>
</table>

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
3. PROBLEM SOLVING – Has a thorough understanding of all dimensions of the operation. Can operate and maintain the equipment efficiently. Is able to analyze and solve problems. Fixes problems completely; doesn’t Band-Aid problems.

4. PRODUCT QUALITY – Is complete and thorough in doing QC checks. Knows the operating and analytical limits to run the system within. Strives to run on target specifications. Understands and follows SPC rules. Uses sensory as a means to correct quality problems. Initiates Hold Orders as appropriate. Knows and follows refeed and blending procedures. Ships quality cases.

5. TEAMWORK – Follows team norms of behavior and confronts those that do not. Contributes to the overall effectiveness of the team. Contributes to team meeting discussions. Takes ownership of “Team Tasks.” Suggests ideas and improvements to help the team work better together. Is candid and honest in interpersonal interactions. Has a high level of respect and integrity. Is flexible and helps others.

6. SAFETY – Follows all Plant operational safe practices. Wears bump cap, safety shoes, hearing protection as appropriate. Does not take risks with his/her or team members personal safety. Keeps work area clean and orderly. Takes an active role in recommending and accomplishing ways to make the plant safer.
7. SANITATION – Completes all sanitation tasks on a daily basis. Follows all plant GMPs. Cleans thoroughly and completely.

8. ATTENDANCE – Has an acceptable attendance record; is reliable and punctual. Keeps team and team leader appropriately informed regarding absences. Schedules PTO in a constructive manner.

9. INTERPERSONAL/COACHING SKILLS – Is even tempered and does not display wide swings of emotion. Is open, constructive and positive while giving and receiving feedback. Brings disagreements or problems out in the open and tries to resolve them. Provides helpful suggestions and effectively trains others. Relates to people openly. Builds working relationships with others. Demonstrates effective listening and communication skills. Provides positive feedback and recognizes exceptional efforts. Respects others opinions and ideas. Is an ardent supporter of plant and coworker development.

10. LEADERSHIP – Actively participates in a committee or coordination role. Attends all required committee meetings. Communicates committee feedback from their team to the committee. Communicates committee information to their team during team time or by other means. Takes on projects or assignments from the committee and completes them in a timely manner. Takes ownership of team and individual tasks. Leads teammates to high performance.
10 September 2005

Tracy Maylett

Protocol #: E0805D08

Project Title: A Study of the Relationship of Multi-Rater Feedback to Traditional Performance Appraisals

Dear Tracy,

Thank you for submitting your application for review to the Pepperdine Graduate and Professional Schools Institutional Review Board (GPS IRB). The IRB appreciates the work that you and your dissertation chairperson, Kent Rhodes have completed on your proposal. Upon review, the Protocol Review Subcommittee in Education of the GPS IRB has determined that the project meets the requirements for exemption from IRB review under the federal regulations that govern the protections of human subjects (45 CFR 46 - http://www.nihtraining.com/ohsrsite/guidelines/45cfr46.html). Specifically,

Category (4) of 45 CFR 46.101. Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

Based upon review, the GPS IRB has determined that your proposed study is exempt from further IRB review.

Please note that your research must be conducted according to the proposal that was submitted to the IRB. If changes to the approved protocol occur, a revised protocol must be reviewed and approved by the IRB before implementation. For any proposed changes in your research protocol, please submit a Request for Modification form to the GPS IRB. Please be aware that changes to your protocol may prevent the research from qualifying for exemption from 45 CFR 46 and require submission of a new IRB application or other materials to the GPS IRB. Because your study is exempt from IRB review, there is no requirement for continuing IRB review of your project.
A goal of the IRB is to prevent negative occurrences during any research study. However, despite our best intent, unforeseen circumstances or events may arise during the research. If an unexpected situation or adverse event happens during your investigation, please notify the GPS IRB as soon as possible. If notified, we will ask for a complete explanation of the event and your response. Other actions also may be required depending on the nature of the event.

Please refer to the protocol number denoted above in all further communication or correspondence related to this approval. Should you have additional questions, please contact me. On behalf of the GPS IRB, I wish you success in this scholarly pursuit.

Sincerely,

Kay Davis, Ed.D.
Member, Protocol Review Committee in Education, GPS IRB

cc: Dr. Lee Kats, Associate Provost for Research
Ms. Ann Kratz, Human Protections Administrator
Dr. Michael Feltner, Chairperson GPS IRB
Dr. Laura Hyatt
Dr. Kent Rhodes, Chairperson