2006

Selecting Leadership: An Analysis of Predictors in Assessing Leadership Potential

Andrea L. Zavakos
Antioch University - PhD Program in Leadership and Change

SELECTING LEADERSHIP:
AN ANALYSIS OF PREDICTORS
IN ASSESSING LEADERSHIP POTENTIAL

ANDREA LYNN ZAVAKOS

A DISSERTATION

Submitted to the Ph.D. in Leadership & Change Program
of Antioch University
in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

September 2006

APPROVAL OF DISSERTATION COMMITTEE

This is to certify that the dissertation entitled:

SELECTING LEADERSHIP: AN ANALYSIS OF PREDICTORS IN ASSESSING LEADERSHIP POTENTIAL

Prepared by:

Andrea L. Zavakos

Is approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Leadership and Change.

Approved by:

Chair - Jon Wergin, Ph.D. Date

Member - Elizabeth Holloway, Ph.D. Date

Member - Carol Baron, Ph.D. Date

Member - Marc Wenzel, Ph.D. Date

External Reader - Dale Rose, Ph.D. Date

DEDICATION

This dissertation is dedicated to my beautiful sons. You are the air I breathe. …And such leadership potential in both of you!

ACKNOWLEDGEMENTS

I would like to acknowledge the following people:

My son, George, who was patient with me and found ways to entertain himself when "Mommy had to do more homework."

To George's father, who always came to the rescue when he was needed.

My son, David Bishop, his wife Melanie, and their new son, Maddox.
You have given me great joy.

My parents. Thank you for always believing in me.

Dr. Peggy Marshall, who talked me into applying for the program, and then throughout the entire experience reached back to prepare me for what was to come. Here's to the new leadership theories we will create together! You are a dear, dear friend.

Dr. Jon Wergin, who could not have been a better committee chair and mentor.

Dr. Elizabeth Holloway, who advised me wisely throughout the program.

Dr. Carol Baron, who gave so generously of her time and energy, and for her clear and logical quantitative brain that never once missed a beat.

Dr. Marc Wenzel, who came through in a pinch several times, delivering research options, networking contacts, and encouragement.

Dr. Richard Couto, who extended his hand in friendship and aid endlessly. You are a key reason this dissertation is in print; and I hope we will always be friends.

Dr. Ken Kramer, a one-of-a-kind comrade and mentor who, among other things, painstakingly edited the documents I submitted for this program without complaint.

Howard and Sally Stevens, gurus and friends extraordinaire.

My fellow students at Antioch: What amazing connections we have made with each other!

The owners of Brower Insurance Agency, whose strong ethics and generous nature convince me every day that there could not be a better employer anywhere.

And finally, to my grandparents - who genetically gifted me with curiosity, chutzpah, and grit.

ABSTRACT

The purpose of this study was to identify predictors of leadership using a newly developed assessment for leadership selection within the healthcare industry by comparing assessment scores to supervisor rankings of the subjects. The study population consisted of 195 employees of 11 different hospitals. Each of the participants completed the Healthcare Leadership Inventory (HLI) assessment; their immediate supervisors completed performance ratings for them. None of the instruments were designed by the researcher. The dependent variable of the study was the supervisor-provided factor of Promotion Potential. Stepwise multiple regression was the main analytical approach.

The analysis yielded two predictors of leadership success from the HLI assessment (Achievement Orientation and Openness to Change) and five from the Supervisor Ratings (Multi-Tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation). The identified predictors from each instrument had construct symmetry, although they were not statistically duplicative. The predictors from Supervisor Ratings provide some insight into the implicit leadership theories shared by management personnel in the healthcare industry. The HLI assessment factors of Achievement, Conscientiousness, Innovative, and Customer Focus had significant correlations with their counterparts from Supervisor Ratings. The Critical Thinking factor surprisingly did not significantly predict leadership potential or correlate with any of the other factors.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION

"In a knowledge economy, companies with the best talent win. And finding, nurturing, and developing that talent should be one of the most important tasks in a corporation. So, why does human resources do such a bad job – and how can we fix it?" (Hammonds, 2005, p. 41). In this introductory statement to his controversial article entitled "Why We Hate HR," Hammonds succinctly describes the importance of marrying human resources practices with scholarly research to improve business results. Improving those business results is especially important in the rapidly changing healthcare industry because "Today's healthcare leaders face crises involving a multitude of complex challenges, such as improving the quality of care in the face of spiraling healthcare costs, overbearing regulations, shortages of skilled healthcare providers, and lack of access to even basic care for millions of Americans" (Altman, 2006, p. 14). A first step toward having the best leadership talent for the healthcare industry is to find ways to better predict candidates' job success. This study offers insight for practitioners who are interested in identifying predictors of leadership success for their organizations.

Background of the Problem

The State of the Art

Human resources and hiring managers need a reliable way to identify and select good leaders – leaders who will successfully drive the initiatives of their companies over the long term. In an attempt to assess leadership potential and ability, practitioners sift through resumes, review data from interviews, criminal background checks, employment and educational verifications, employment references, 360º reviews, psychological assessments, financial results, assessment center results, performance appraisal scores, and any other information they feel might be relevant. Others, confused and overwhelmed, rely solely on intuition and "personal

chemistry" to make hiring decisions. Often less aware of the major leadership theories and their efficacy than their more scholarly counterparts, practitioners are searching for the proverbial needle in a haystack when selecting leaders for hire or promotion. Interviewing is heavily relied upon in most corporations for determining fit for a job; and traditional unstructured interviewing can be ineffective. The more structure and standardization that can be added, the more predictive the employment interview will be for job success. In fact, Williamson *et al.* (1997) cited a number of meta-analyses with validities for structured interviews ranging from .24 to .34, compared with much lower validities for unstructured interviews ranging from .11 to .18. Schmidt and Hunter (1997) conducted a meta-analysis in which they cited validities for structured interviews to be .51, much stronger than the Williamson study. Whatever the statistics one chooses to site, a general perception in the professional human resources arena is that, when selecting someone for a leadership position, a hiring manager would do just as well by flipping a coin as opposed to relying solely on an interview process that is not fortified with a structured, behavior-based approach.

Knowing "what to look for" is the first challenge for practitioners. They struggle with identifying the traits, qualities, skills, behaviors, habits, attitudes and motivations that might predict which leaders can be successful in their organizations. Even the scholars do not agree – and tend to publish conflicting or confusing information on what constitutes a good leader. Practitioners often perceive the scholarly research to be lofty, disconnected from the "real business world," and unrelated to actual business results. They may not recognize it, but many practitioners are reacting to the dilemma between rigor and relevance (Schön, 1983), and most times are choosing relevance. Schön pointed out that there is a hierarchical separation of research

versus practice, and that research tends to be seen as the superior activity. However, in the business world, results are paramount.

Even if practitioners do find a theory of leadership to which they and their organizations can subscribe, the theory is seldom connected with any sort of instrument that practitioners can actually use in the selection process. So the second major obstacle for practitioners, even when they are able to process the ever-expanding body of knowledge about leadership and a hopefully viable approach, is knowing "how to look for it." Most do not have a practical and systematized way to apply leadership theory to their selection processes.

The Healthcare Industry

The healthcare industry is an extremely unique and difficult environment in which to operate as a hiring manager, human resources practitioner, or leader. "The workforces of hospitals are among the most highly educated in the service sector; however, the nature of professional education makes for a very complicated tapestry of relationships" (Garman & Tran, 2006, p. 152). Human resources professionals know that one way to measure a position's importance for compensation purposes is to examine the seriousness of impact of decisions made by the position incumbent. Considering the fact that decisions made within a healthcare context can literally determine someone's life or death, the distinctive challenges healthcare leaders in particular face become more apparent. Because many "Leadership positions in a broad range of health care organizations – including those in the pharmaceutical, biotechnology, and health insurance industries as well as hospitals and other provider systems – are filled by scientists and physicians" (Eiser, Eiser, & Parmer, 2006, p. 3), there are additional leadership challenges in the healthcare industry. Scientists and physicians *do not* often receive formal training on leadership and management because they are focusing on their scientific specialties. There are often great

differences in the "goals, organizational power, and income" (Garman & Tran, 2006, p. 152) of physicians as compared to their non-physician peer leaders, which requires even more complex interpersonal skills than the average leader to manage correctly. These internal and external forces test healthcare leaders in a one-of-a-kind way, and must be addressed and selected for in specialized ways as well.

How to Improve Leadership Selection Decisions

So, how does one help healthcare practitioners better identify and select leaders for their organizations? First, one must examine and improve the methods of inquiry used in making selection decisions. Second, a better conduit must be developed between theory and the practical world, so that theoretical information is easier to access, understand, and apply.

Adapting Inquiry Methods

In the world of recruiting and selection regardless of industry, there is a large "store" of available means to assess candidates. Although companies and hiring managers frequently differ on how to weight the importance of specific assessment tools, the following elements are generally accepted approaches to hiring (once the position has been accurately described and applicants have been generated), with some used more or less than others (Boulden, 2002; Cascio, 1989; Dessler, 2003). The elements which follow are also represented in Figure 1 for visually-oriented readers.

Reviewing resumes/applications: The HR professional or hiring manager screens the information provided to determine if the candidate meets minimum requirements for the position. A popular approach is to use "yes," "no," and "maybe" piles for incoming resumes. Biographical data measures have predictive validity of r = .35 (Schmidt & Hunter, 1998).

Telephone screening: HR or the hiring manager contacts the candidate to ask a limited number of questions, such as if still interested in the position, salary range, and willingness to relocate. This process usually results in a "pass/fail" score.

Interviewing: Hiring decision-makers meet with the candidate in person, via telephone, or other means to better determine fit. Interviews can range from conversational meetings with no real direction to behavior-based approaches and grading systems. The more objective the interview can be made (through structure, focus on the characteristics needed for the position, etc.), the more it can be relied upon to actually predict success in the position (Huffcutt, 1994, p. 190). The interview process is interactive, with each party gleaning information about fit, and affecting the other party's view of the same. Unstructured interviews are estimated to have predictive validity of r = .38 while structured interviews enjoy a much higher validity of r = .51 (Schmidt & Hunter, 1998).

Realistic job previews: The candidate is provided with a realistic – and often hands-on – view of the position in an actual or simulated work environment. A realistic job preview will yield a pass/fail result – the candidate will either react negatively or positively to the environment and may self-select at that point not to continue in the selection process. Job tryout procedures have an estimated predictive validity of r = .44 (Schmidt & Hunter, 1998).

Testing and Assessments: The candidates' fit for the position is tested in one or more of the following categories: skills, personality, intelligence/cognitive ability, vocational interest, or assessment centers. Most assessments provide some sort of ordinal (scored) result. The predictive validity of testing and assessments ranges from r = .31 for conscientiousness tests to r = .48 for job knowledge tests; and integrity tests have predictive validity of r=.41 (Schmidt & Hunter, 1998).

Work Simulations: The candidate completes a work assignment that is similar to or identical to an activity that would be required in the position. Predictive validity for these ranges from r = .36 for assessment centers to r = .54 for work sample tests (Schmidt & Hunter, 1998).

Reference checking: Hiring decision-makers speak with people who have some level of experience with the candidate. Reference checks can range from undirected conversations with personal friends to behavior-based, focused sessions with the candidates' professional contacts. Similar to interviewing, the more objective reference checks can be made (talking with past supervisors, adding structure, focusing on the characteristics needed for the position, etc.), the more they can be relied upon to actually predict candidate success in the position. The reference check process is interactive; with each party gleaning information about fit, and affecting the other party's view of the same. Most reference providers will feed back their impressions about the potential employer to the job candidate based upon their experience during the reference check, and this affects the candidate's view of the company and position. The predictive validity of references checks is r=.26 (Schmidt & Hunter, 1998).

Employment verification: Past employment and related data, such as reason for termination provided by the candidate are confirmed. Results are usually pass/fail. Job experience itself has a predictive validity of .18 (Schmidt & Hunter, 1998), and verifying the accuracy of employees' claims of past employment is important as well, considering a recent report from ADP Screening and Selection Services asserting that 44 percent of applicants misrepresented their work history, as found during their performance of 2.6 million background checks in 2001 (Babcock, 2003).

Educational verification: Information provided by candidate regarding educational background, degrees and grade point averages is substantiated. Results are usually pass/fail (either the candidate has the degree cited or not). Years of education have a predictive value of r=.10

(Schmidt & Hunter, 1998), and verifying that the education or degree represented was actually attained becomes increasingly important in light of recent estimates that 41 percent of job candidates have lied about their education on their resume (Babcock, 2003).

Background checks: Criminal history, credit reports, and other relevant information about job candidates is evaluated.

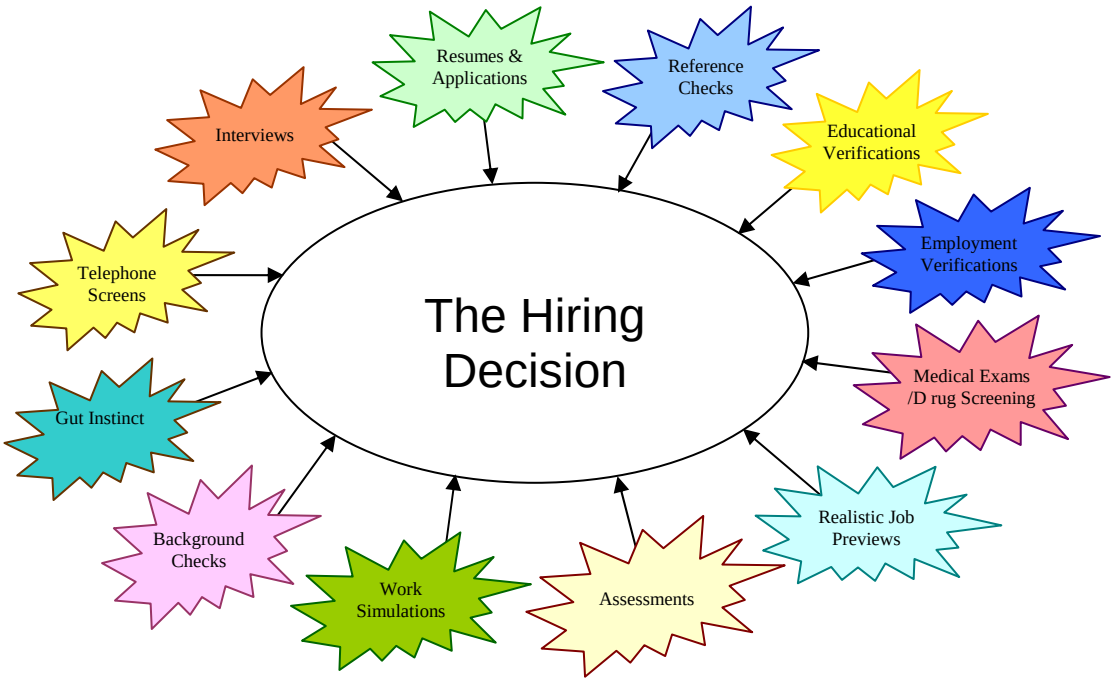
Medical examination: A medical professional who understands the job requirements and essential functions assesses the candidate for fit. Results are usually pass/fail (either the physician believes the candidate can meet the physical requirements of the position or not). While a medical examination is normally conducted after an offer of employment is made, it can result in the elimination of the job applicant.

Drug screening: A laboratory verifies that the candidate's blood or urine does not contain unacceptable levels of certain substances. Results are usually pass/fail (either drug levels are over the acceptable level or not). While drug screenings are normally conducted after an offer of employment is made, they can and do result in the elimination of job applicants.

Gut instinct: Using tacit knowledge (intuition) in decision-making about a job candidate. Technically speaking, "gut feeling" is less a selection method and more the filter through which the results of all the other selection techniques is passed. However, its importance in the selection process is so paramount that it warrants a dedicated heading.

Figure 1.

The Hiring Decision



Research Methods in Selection

The process of answering a research question compares quite readily with the selection process. The researcher or hiring manager decides what information is needed to make certain decisions, how that information will be obtained, and how it will be used. So, using this analogy between making a hiring decision and answering a research question, one can look at the aforementioned items as the various methods of inquiry (some more scientific than others). These methods of inquiry could then be categorized into three distinct groups, based on their epistemological bases:

Quantitative: These elements of the decision-making process are based solely (or at least largely) on objective measurable data. This category would include telephone screening, testing and assessments, past employment verification, educational verification, medical examination, and drug screening. The "answers" to the inquiries in this category are in some form of "pass/fail." For instance, the educational verification either confirms the candidate has the degree the candidate claims to have or does not. Likewise, the candidate either passes or fails the drug screening. Assessments, too, would fall into the quantitative category because they will provide scoring or other indicators that relate to fit for the position or some other relevant variable.

Qualitative: This includes those elements of the decision-making process that involve some interpretation of personally interactive exchanges, such as interviews and reference checks.

Tacit: These elements involve one's intuition or "gut feel" (Sternberg & Horvath, 1999). It is important to note that this intuitive approach can inform or skew the preceding two categories. It is easy to see that one's gut feel about a candidate could affect his/her view of how the interview or reference checking went. It may be more difficult to see how tacit knowledge can affect quantitative data, but it can easily happen with assessments and testing. The hiring manager may have selected a criterion, an assessment, or series of assessments based on his/her gut feel (or perception of "face validity") regarding what the job requires, as opposed to having done a rigorous review of the requirements of the position in consultation with experts in assessments. In discussing this issue, Heneman (1980) noted, "…there is the possibility that the decision-maker may interpret, or even ignore, valid information about applicants" (p. 56).

Malcolm Gladwell (2005) addressed this phenomenon is his book, *Blink*, by discussing the large percentage (58 percent) of Chief Executive Officers (CEOs) of Fortune 500 companies in the United States who are six feet tall or taller, as compared to the much smaller percentage

(14.5 percent) of the general male population in the United States who are in the same height range. “We have a sense of what a leader is supposed to look like and that stereotype is so powerful that when someone fits it, we simply become blind to other considerations” (p. 140).

The Pygmalion Effect

The “Pygmalion effect,” Merton’s (1957) self-fulfilling prophecy theory, provides us with another lens into how tacit and other less explicit forms of knowledge can affect perceived or real performance in individuals, which in turn affects our decision-making about them. This phenomenon, also called the expectancy effect, is often described by an experiment conducted by Rosenthal and Jacobson in which teachers were told that certain children, who were actually selected at random, had been categorized as “blooming” and were expected to experience dramatic gains in cognitive abilities in the upcoming school year (Rosenthal & Jacobson, 1968). After eight months, the students were re-tested; and those identified as bloomers in fact showed greater intellectual gains that the other students. Dr. Rosenthal hypothesized that because the teachers communicated their high expectations to, and had more confidence in the bloomers, those students progressed more favorably. Is this phenomenon repeating in the corporate world with leadership selection and development programs? Research has repeatedly shown that employees’ performance improves when their supervisors express positive expectations and confidence in their abilities (Davidson & Eden, 2000; Kierein, 2000; Rheem, 1995; Rosenthal, 1997; Sutton & Woodman, 1989). If age-old ideas about leaders are myths, do they become self-fulfilling prophecies because decision makers seek out people who fit the potentially inaccurate paradigms they hold?

Alternatively, the “Golem effect” is described as the corresponding decrease in productivity that occurs when the supervisor has negative views about the employee’s abilities

(Davidson & Eden, 2000). Davidson discusses the fact that disadvantaged women have suffered negative consequences in the past that can be attributed to the Golem effect. Females have not only been overlooked for leadership positions due to their gender, even to the extent that they have been encouraged to behave differently. “In fact, popular literature is filled with self-help techniques to enable women to overcome the inherent defects that result from being female” (Heilman, 1997, p. 878). Is it possible that, like disadvantaged women who *do not* fit the “traditional” leadership traits, others - male or female - who have promising leadership potential, but who *do not* fit the “great man” leadership trait schema, are being dismissed by corporate America?

The Research Question

So, again analogizing the hiring/promotion decision to a research question, there are three main inquiry approaches to making selection decisions: quantitative, qualitative, and tacit. Researchers in the field (Lievens, Highhouse, & Corte, 2005; Philbrick, Bart, Sparks, & Hass, 1999; Williamson, Campion, Malos, Roehling, & Campion, 1997) suggest that reliance on any one selection data point to the neglect of others is a mistake, and that an appropriate reliance on a variety of approaches is best. Upon having a clear understanding of the three methods of inquiry in the selection process, practitioners can then work to improve the methods they use within each category. For instance, a great number of tools, templates, and training programs are available to teach practitioners how to conduct behavior-based or other types of interviews that yield more predictive results than an unstructured, conversational interview.

Certain selection approaches cannot be improved upon a great deal, such as criminal background checks or educational verifications. These tend to be nominal – pass and fail – categories. Aside from verifying that data received is accurate, clarifying the acceptable result

and holding to it, using good judgment as needed, and making the process more efficient, there is not really any approach for criminal background screens for instance, that, if applied, would yield better information for the employer to use in the hiring process. However, improving other quantitative approaches, particularly the use of assessments, could substantially increase the accuracy of hiring decisions.

Reliance on Tacit Knowledge

A chasm currently exists between the way scholars and practitioners view leadership. The fact is that most of the primary research is simply not easy to read. Ford states that management research should, among other things, "be coherently written" (2005, p. 33). Those who have not learned or practiced the reading of correlation tables, for instance, would find it difficult to quickly interpret their meaning. Unless someone bridges the gap and puts the information into layperson's terms in a trade magazine such as *HR Magazine*, most working professionals will not see it. Charles O'Reilly is a Professor of Human Resources and Organization Behavior, and Jeffery Pfeffer is a Professor of Organizational Behavior, both at the Graduate School of Business at Stanford University. They have done what many academics are unable to do - continued to produce quality research *and* translate it into digestible format for the non-academic reader (O'Reilly & Pfeffer, 2000). For instance, they have found practitioner-oriented forums and formats to interpret research for practitioners on the effects of organizational culture on individual performance (2000), Cisco Systems' secret of success (2000), and employee turnover in hospitals (1987). Rose and Fiore have also helped to bridge the gap by focusing on practical ways to evaluate human resources programs when the "high fidelity scientific model" (p. 236) is not feasible (1999).

The sheer volume of the scholarly literature is another complicating factor. Should a practitioner decide to take a more scholarly approach to his practice, where does he begin? The scholars offer far too much information for most working professionals to review and synthesize into a viable theory of leadership in their spare time. As Ford states, “The culture and competitive environment of business enterprises require that managers make decisions fast and often with insufficient information…Researchers are trained to search for information even at the risk and expense of taking more time” (2005, p. 32). To further this point, Dennis Ahlburg (1992) surveyed a group of recruiting professionals and a group of students, and found that “US personnel neither knew nor used the research base of their profession...”(p. 467). Further, he stated, “It seems that practitioners do not believe the research evidence or are prevented from using it by established practices in their organizations, or they consider other issues more important than validity in the choice of methods for selecting managers” (p. 467).

Ford *et al.* (2005) asserted that one of the reasons many practitioners have little interest in leadership research is that they do not view the research as relevant to their work. Additionally, they stated that scholars view practitioners as simply being interested in quick fixes with the latest management fads. They discuss a number of ways to enhance the likelihood of leadership researchers producing practice-relevant studies such as adding a “Why This Research Should Matter to Managers” (p. 35) section at the beginning of research articles, and having managers catalog and provide their recurring issues to researchers. They provide a simple yet descriptive explanation of this process: “translating management thinking into management action” (p. 35). Another way of describing this phenomenon is “praxis,” which means the “practical application of a branch of learning” (*The American Heritage Dictionary*, 2001, p. 661). It appears that Ford *et al.* support the case for praxis in the selection of leaders (2005).

So, how are managers and HR people making decisions about what to test for or train for at work? Having worked in the field of human resources for more than 18 years, this researcher's experience is that gut instinct is a strong influence. This is not to suggest that managers and HR people rely solely upon intuition, as "There are many examples of how managers have accepted and benefited from management research" (Ford *et al.*, 2005, p. 25). But it is not difficult to see how the scenario unfolds: a busy manager sees a summary of an assessment and finds face validity in the description – it appeals to him in some way. Even if a manager wants to use an objective assessment to make a hiring decision for a subordinate manager, the assessment itself is often chosen with little factual or theoretical basis – she might simply have an affinity for the description of the assessment. Furthermore, when a favored candidate's assessment results have negative results, managers may even then discount the same assessment tool they originally selected, because it goes against their gut instinct – their intuitive sense of the candidate's fit (Heneman III, Hamstra, & Brown, 1980). Even if an HR practitioner with a solid personal theory of leadership decides to take a more well-rounded approach to selection processes, the next obstacle to overcome is in knowing how to relate this theory of leadership to candidate selection. Finding predictors for leadership in the form of an assessment is dubious at best because there is no clear, singular, and enduring definition of what leadership is.

Leadership Defined

How have leaders been recognized, identified, or assessed over the years? The first step has been in attempting to define leadership in various ways: who the leader is, what the leader does, how the leader interacts with followers, and the environment of leadership. Confucius, Lao Tzu, Plato, Aristotle, Plutarch, and Machiavelli were all concerned with the characteristics of effective leaders, and in Plato's case their selection. For instance, while Confucius helped prepare

many men for political leadership, he warned against the art of rhetoric (Confucius & Slingerland, 2003). Plato believed leaders should be well-educated in order to lead well (Plato & Grube, 1974). Aristotle felt strongly that leaders should have moral goodness (Aristotle & Burnet, 1973). Machiavelli's *Prince* emphasized that the successful prince must exhibit prowess in both favorable and adverse circumstances (Machiavelli & Bull, 2005). In more modern times, researchers and practitioners continue to attempt to define leadership from seemingly every possible angle. Correspondingly, both the field of scholarly leadership study and the business world have been flooded with assessments and other methods and instruments designed to predict whether a particular individual can be successful as a leader. However, in the end, one cannot determine what good leadership is without also looking at the context in which leadership is occurring or is needed. There have been some limited attempts to describe leadership in context and to relate organizational culture to leadership efficacy, but nothing that has linked culture and leadership attributes in a way that allows practitioners to use the linkage to select good leaders for their particular organizations.

Purpose of the Study

The purpose of the study was to examine the degree to which a new healthcare leadership assessment instrument relates to potential for promotion ratings provided by supervisors. The hope was that this investigation would reveal specific indicators of leadership potential for various organizational contexts within the healthcare industry. This study provides initial evidence of a cluster of leadership indicators which can be used in leadership selection and development for the healthcare industry. Leadership selection error can be expensive - even fatal - to organizations. When companies have a reliable tool that helps them more effectively predict

leadership potential and make better hiring decisions, they will mitigate risk and experience more overall business success.

Summary

The selection of leaders has long occupied human attention. While human resources professionals do employ useful tools such as assessment centers, personality tests, and structured interviews, a disparity continues to exist between theory and practice in leadership selection. The researcher hopes that this dissertation will help to bridge the gap by assessing the effectiveness of a new leadership instrument in the healthcare industry.

Organization of the Study

This dissertation consists of five chapters. The first chapter establishes the basis and rationale for the current study. Chapter Two presents relevant literature regarding leadership and sets forth a comprehensive picture of its definition and means of measurement including the predominant leadership assessments, their efficacy, and the theories on which they are based. Chapter Two also addresses areas in which earlier works may be incomplete and implications for future research. Chapter Three focuses on the quantitative design of the study, and the procedures used to conduct it. Results and findings of this dissertation are presented in the fourth chapter, with conclusions and recommendations provided in Chapter Five.

CHAPTER II – REVIEW OF THE LITERATURE

Introduction

In Chapter One, the uniqueness of the healthcare industry, the elements of selection, and the gap between scholarly research and practitioner application of selection best practices were reviewed. In this chapter, the literature cited will provide a closer look at leadership theory and the practice and importance of leadership selection.

Businesses in the United States spend an enormous amount of time, energy, and money in efforts to identify and predict effective leadership – totaling at least 40 billion dollars a year (McCall, Lombardo, & Morrison, 1988). Yet selection error continues to be a problem in this country and elsewhere. One merely has to turn on the evening news to learn about the latest scandals in leadership at companies such as Boeing, Tyco, and Enron. These high-profile leadership debacles have been most closely associated with ethical issues. Kellerman (2004) hypothesized that leaders who do not live up to expectations are either incompetent or evil.

However, many less public leadership failures can be attributed to an improper "fit" within the work environment. Edgar Schein (1978) was a forerunner in this concept, exploring how organizations and employees can meet each other's needs to enhance organizational performance and survival. He described a matching process that should occur between an organization and its people in such areas as recruiting and selection, training and development, career counseling, and organizational rewards. Goffee and Jones (1998) described four types of organizational culture along with their "evil twins," and provided insight into how to survive in each. A year later, Schein (1999) provided another perspective of culture typologies and survival techniques. He defined culture as "the sum total of *all* the shared, taken-for-granted assumptions that a group has learned throughout its history" (p. 29).

Selection errors, whether due to incompetence, evil, or simply a poor fit in terms of culture or skill set, can have disastrous results. Researchers have studied good leadership, leadership derailment, and a host of other potential predictors of leadership success or failure. However, most practitioners do not apply the results of academic research within their organizations. Buckley *et al.* (1998) administered a five-item questionnaire to 113 human resources managers while at a conference, asking whether they were familiar with current academic research, how applicable the research was to real-world problems, and to what extent they have used academic research to solve problems in their organizations. The results indicated that practitioners are aware of academic research, but do not believe researchers understand the business world or are focusing on the right problems. Another study involved 959 human resources professionals, with similar findings:

> In particular, practitioners place far less faith in intelligence and personality tests as predictors of employee performance than HR research would recommend. Practitioners are somewhat more likely to agree with research findings when they are at higher organizational levels, have SPHR [Senior Professional in Human Resources] certification, and read the academic literature (Rynes, Colbert, & Brown, 2002, p. 149).

If the disparities between scholarship and practice can be minimized, then inasmuch as selection error can be reduced, decision makers can incorporate that knowledge into practice within their organizations. To that end, a detailed review of the theoretical history of leadership, personnel selection, and predictive assessments is in order.

Theories of Leadership

Historical Perspective on Leadership Theory

Rost provided an exhaustive literature review detailing the history of thinking about leadership (1991). He found the first definition of leadership in Samuel Johnson's 1755 dictionary in which the word "leader" was defined as "one who goes first." Despite a number of forerunners such as Mr. Johnson, "the systematic social scientific study of leadership did not begin until the early 1930s" and predominantly reflects Western culture (House & Aditya, 1997, p. 409).

Trait Theory

The Greek philosopher Plato (circa 428-c. 347 BC) may have offered one of the earliest formal insights into the trait theory by identifying certain qualities he felt were essential to political leaders (Plato & Grube, 1974); this topic received much more attention centuries later. In the 1930s, scholars began to systematically focus on individual characteristics as indicators of leadership ability. Bass (1981) offers an efficient summary of the trait research and theories from this period in history:

> Until the 1940s, most research about leaders and leadership focused on the individual traits of consequence. Leaders were seen to be different in various attributes and tested personality traits than were non-leaders. Two questions were posed: What traits distinguish leaders from other people? What is the extent of those differences? Then pure trait theory fell into disfavor. Stogdill's (1948) critique concluded that both person and situation had to be included to explain the emergence of leadership (p. 38).

Bass concluded this discussion with a hint that trait theory is still alive and well. Let us examine some other views on trait theory.

Gary Yukl stated:

> The early leadership researchers were confident that the traits essential for leadership effectiveness could be identified by empirical research. The kinds of traits studied most often in the early leadership research included physical characteristics (e.g., height, appearance), aspects of personality (e.g., self-esteem, dominance, emotional stability), and aptitudes (general intelligence, verbal fluency, creativity). Many of the early studies compared leaders to non-leaders or examined the attributes of emergent leaders in newly formed groups (2002, p. 177).

Northouse summarized trait theory very well:

> The trait approach does not lay out a set of hypotheses or principles about what kind of leader is needed in a certain situation or what a leader should do, given a particular set of circumstances. Rather, this approach emphasizes that having a leader with a certain set of traits is crucial to having effective leadership. It is the leader and his or her personality that is central to the leadership process" (1997, p. 21).

Intelligence

The trait of General Mental Ability (GMA) has been explored extensively since the early twentieth century (Spearman, 1904). A recent meta-analysis showed a computed predictive validity of r=.51 for general mental ability tests, which is on the high end as compared to integrity tests (r=.41), conscientiousness tests (r=.31), and second only to work sample tests at r=.54 (Schmidt & Hunter, 1998). When coupled with personality-related predictors, intelligence-oriented assessments have shown to be quite useful in predicting leadership. Lord, de Vader, Alliger (1986) explored the relationship between personality traits and leader emergence in their meta-analysis, and found intelligence (among other things) was significantly related to leadership

perceptions. Vecchio (1990) found that leadership intelligence was highly correlated with group performance. Judge, Colbert, Ilies (2004) also examined the relationship between leadership and intelligence and found that the relationship between intelligence and leadership was not as strong as previously thought. In that same year, however, Schmidt and Hunter's (2004) study comparing general mental ability with job performance was published, stating that cognitive ability tests are able to predict occupational performance "better than any other ability, trait, or disposition, and better than job experience" (p. 162). Although the debate on the predictive value of GMA continues, it does hold face validity because:

> If one worker learns faster than another, the same amount of experience will produce a higher level of performance in the fast learner than in the slow learner. It is GMA that turns experience into increased job knowledge and hence higher performance (Schmidt & Hunter, 2004, p. 167).

The "Big Five" Personality Theory

Because so many leadership assessment instruments contain some aspect of the Big Five Personality Theory (Conway & Peneno, 1999; McCormack & Mellor, 2002; Morrison Jr., Abraham, & Dennis, 2004; Rubenzer, Faschingbauer, & Ones, 2000; Stricker & Rock, 1998; Tanoff & Barlow, 2002; Wielkiewicz, 2002), it is important to specifically review it understanding that it is a subset of Trait Theory. Although the Big Five Personality Theory is not specifically or exclusively associated with leadership *per se*, a great deal of research has gone into explaining leadership through personality traits (Lord, de Vader, & Alliger, 1986), as discussed in the previous section. Although the focus on personality traits fell out of favor by the early 1950s, recent research has encouraged another look (Atwater & Yammarino, 1993).

Goldberg (1990) described the process through which the Five Factor Model (FFM) was derived, beginning with Raymond Cattell, who identified more than 12 factors which were then reduced to five using orthogonal rotational methods. These factors are Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness. Schmidt and Hunter (2004) place specific emphasis on the Conscientiousness factor, stating, "In the prediction of performance on the job, only of the Big Five traits – Conscientiousness – has been found in the meta-analytic studies to function like GMA in that it consistently predicts job performance in all job families studied (Barrick & Mount, 1991; Mount & Barrick, 1995)" (2004, p. 169). Proponents assert that the five factors are stable over time and are consistent personality factors across all situations and cultures (McCrae & Costa Jr., 1997), although it is widely acknowledged that "the universal existence of these FFM traits does not signify these play the same role across cultures (Bond and Forgas, 1984; McCrae and Costa, 1997)" (Leung & Bozionelos, 2003, p. 63). The International Personality Item Pool (IPIP) developed by Goldberg (1990) is often cited in association with the FFM, and has recently been shortened and made available for Internet use (Buchanan, Johnson, & Goldberg, 2005). But is the FFM the "silver bullet" of leadership prediction? "The Five Factor Model is useful in discussing important recurrent features of personality, but it is incomplete" (Braun, Jackson, Wiley, & Messick, 2002, p. 83). In other words, the FFM may be a necessary but not sufficient component in assessing for leadership potential.

Attempts by researchers to understand how personality manifests itself in leadership success continue, and may be an important key to improving leadership selection practices. However, the notion that there is one "silver bullet" or specific trait that every good leader should have is simply not practical in this complex world. While Trait Theory's importance has

seemed to diminish over time, it is clear that, even today, researchers continue to explore its validity and usefulness. But by the early 1950s, the focus of scholars had shifted to looking at the group versus the individual.

Focus on the Group

Rost described various group-focused definitions of leadership that began to emerge in the 1950s, and quotes Cartwright and Zander's *Group Dynamics* (1953) definition of leadership:

> Leadership is viewed as the performance of those acts which help the group achieve its objectives. Such acts may be termed *group* functions. More specifically, leadership consists of such actions by group members as those which aid in setting group goals, moving the group toward its goals, improving the quality of interactions among the members, building the cohesiveness of the group, or making resources available to the group. In principle, leadership may be performance by one or many members of the group (1991, p. 51).

Using a goal-setting approach in the evaluation of leadership continues the path toward leadership accountability and measurement. One criticism of even the modern literature and empirical studies is that there is not enough of a focus on actual business measures for it to be useful to practitioners. However, scholars began the dialogue about effectiveness many years ago, as referenced in Rost (1991, pp. 52-53):

> A third theme of leadership definitions in the 1950s emphasized effectiveness. Stogdill opened the decade with such a definition: "Leadership may be considered as the process (act) of influencing the activities of an organized group in its efforts towards goal setting and goal achievement" (1950/1958, p. 33). Cattell (1951) defined a leader as a person who has a demonstrable influence on group syntality and stated that leadership is "the

magnitude of the syntality change produced by that person" (p. 175). Syntality is a measure of the group's effectiveness as a group, so Cattell ended up defining leadership by the magnitude of the change in group effectiveness.

Leadership Style

Northouse contrasted the style approach with Trait Theory by pointing out that while Trait Theory focuses on the characteristics of the leader, the style approach has the leader's *behavior* as its focal point. "The style approach focuses exclusively on what leaders *do* and how they *act*" (1997, p. 35).

Some of the most important work relating to style was conducted at Ohio State University, where the researchers used subordinate questionnaires to target leadership behavioral descriptions. This work gave birth to the Leader Behavior Description Questionnaire (LBDQ) that is mentioned elsewhere in this document, and which was later modified into a shorter version by Stogdill that focused on two main dimensions of the original instrument: initiating structure and consideration (Northouse, 1997).

The University of Michigan Studies gave "special attention to the impact of leaders' behaviors," and identified *employee orientation* and *production orientation* as two important dimensions to consider when assessing leadership potential (Northouse, 1997, p. 37). Employee orientation (nearly identical to Ohio State's "Consideration" dimension) results from an emphasis on human relations, while production orientation focuses on technical and production-related aspects of getting the job done. The researchers saw these orientations as each occupying opposite ends of the same continuum, and felt that those leaders who had a strong employee orientation did not have a production focus. However, the researchers at Ohio State felt

differently; and eventually, the researchers agreed that these were two independent dimensions of leadership style.

Northouse spotlighted the development of Blake and Mouton's Managerial Grid, which was later renamed the "Leadership Grid," and which set "concern for people" and "concern for production" to X and Y axes (Blake & Mouton, 1964). The grid identified five types of leadership styles: Authority Compliance, Country Club Management, Impoverished Management, Middle of the Road Management, and Team Management.

During this time, theorists continued their attempts to define leadership. Rost (1991) provides Fiedler's 1967 definition: "By leadership behavior we generally mean the particular acts in which a leader engages in the course of directing and coordinating the work of group members" (p. 56).

Contingency Theory

Theorists began to consider situational variables in the mid- to late-1960s, and developed theories to account for them. Fred Fiedler is considered the grandfather of Contingency Theory; and his (1967) book carried an important title, *A Theory of Leadership Effectiveness*. Not only is it important that researchers were thinking about how situational variables might affect leadership, but they were also beginning to understand the importance of *effective* leadership, and potential measures for this. Yukl summarized this shift in thinking below:

> Comparative research on the way managerial behavior varies across situations…provides some useful insights, but it is only an indirect approach for discovering what type of leadership is optimal in a given situation. A more direct approach is to determine how leaders' traits or behaviors are related to indicators of leadership effectiveness in different situations (2002, p. 208).

Contingency Theory is supported by a great deal of empirical research (Ayman & Chemers, 1991; Baril, Ayman, & Palmiter, 1994; Fox, Hill, & Guertin, 1973; Green, Nebeker, & Boni, 1976; Kennedy, K, & Jr, 1982; Peters, Hartke, & Pohlmann, 1985; C. A. Schriesheim, Tepper, & Tetrault, 1994; Weissenberg & Gruenfeld, 1966), and has broadened our understanding of the importance of situational variables and picking the right leaders for different challenges. Importantly, it is found to be somewhat predictive of leadership success. However, this theory has also received wide criticism due to some level of mysticism about how it works, because the Least Preferred Co-Worker (LPC) scale has questionable face validity, and it does not correlate well with other leadership measures. Contingency Theory can also be difficult to apply due to its complexity (Northouse, 1997).

Path-Goal Theory

In the early 1970s, the Path-Goal Theory became popular. It is based upon the notion that leaders can help followers along their paths by providing them with those things they need to do so. Specifically, theorists began delving further into a matter of great importance: how leadership behavior affects subordinates' achievement of goals. "A motivation theory called 'expectancy theory' (Georgopoulos, Mahoney, & Jones, 1957; Vroom 1964) is used to explain how a leader can influence subordinate satisfaction and effort" (Yukl, 2002, p. 212). Expectancy theory "suggests that subordinates will be motivated if they think they are capable of performing their work, if they believe their efforts will result in certain outcome, and if they believe that the payoffs for doing their work are worthwhile" (Northouse, 1997, p. 89). Path-Goal Theory has evolved to now contain four descriptors of leader behaviors: supportive, directive, participative, and achievement-oriented. Yukl tells us that empirical research has yielded inconclusive results

about Path-Goal Theory's efficacy. The theory has also been criticized because it is based upon the expectancy theory, which does not fully consider leadership situations.

Charismatic Leadership

While some scholars were exploring Path-Goal Theory in the late 1940s, Max Weber was generating ideas that contributed to the concept of a charismatic leader. Weber believed that charismatic leaders arise in times of crisis with revolutionary new ideas that can potentially solve problems, and attract people to help fulfill those visions (Yukl, 2002).

Yukl (2002) described House's explanation of charismatic leadership in 1977. The explanation was based on a set of "observable processes" that could be tested. House helped emphasize the component of *measuring* leadership ability. He claimed charismatic leaders display certain personality characteristics such as dominance, "having a strong desire to influence others, being self confident, and having a strong sense of one's own moral values" (Northouse, 1997, p. 134). The concept of charismatic leadership is often grouped with Transformational Theory, as discussed in the next section.

Transformational Leadership Theory

Northouse (1997) states that the term "transformational leadership" was first used by Downton in 1973, but James McGregor Burns made it famous. In his (1978) book, *Leadership*, Burns set forth two distinct types of leadership: transactional (focusing on exchanges) and transformational (focuses on a connection with others that increases motivation). Bass (1985) expanded upon Burns' ideas, and also created the Multi-Factor Leadership Questionnaire (MLQ) based on Transformational Leadership Theory, which is still used and tested extensively in empirical research on leadership (Avolio, Bass, & Jung, 1999; Bass, Avolio, & Goodheim, 1987; Bycio, Hackett, & Allen, 1995; Goodwin, Wofford, & Whittington, 2001). Transformational

behaviors used as factors in the MLQ are: idealized influence, individualized consideration, inspirational motivation, and intellectual stimulation. Transactional behaviors are: contingent reward, active management-by-exception, and passive management-by-exception.

Yukl stated that "a wide variety of different research methods have been employed in the research on charismatic and transformational leadership. Most of the research has been focused on leader behavior and how it is related to follower motivation and performance" (2002, p. 255). In field study, researchers have found some positive results, at least for some of the factors within the MLQ. For instance, Goodwin *et al.* (2001) found support in their research for their hypothesis that contingent rewards are linked to transformational leadership in their research. Additionally, Bycio *et al.* (1995) conducted a confirmatory factor analysis to compare the MLQ to five facets of transactional and transformational leadership, and found statistical support for the "intent to leave" and "organizational commitment" facets.

Substitutes for Leadership Theory (SLT)

The SLT focuses more closely on measurable leadership outcomes, spotlighting the variables that make leadership either impossible or unnecessary (Bass & Stogdill, 1990). For example, group processes (substitutes) or reward systems that work independently of the leader (neutralizers) can substitute for leadership. Kerr and Jermier (1978) developed this theory, and set out to describe leadership effectiveness through subordinates' commitment to their organizations.

"Role clarity" and "task motivation" are concepts often associated with this model, and are used frequently in the empirical research. Yukl (2002) stated that the "Howell *et al.* (1990) contention that some situations have so many neutralizers that it is difficult or impossible for any leader to succeed" (219). This important concept begins to create a new framework around our

thinking about leaders. So-called "good" leaders may not necessarily be good leaders in all situations. The notion that a good leader may not "have what it takes" to master every leadership situation supports the Contingency Theory, yet conflicts with the Trait theorists' contention that there are certain traits that all effective leaders have and that success or failure is determined by those traits.

Empirical research has yielded some support for the SLT (Alban Metcalfe & Alimo Metcalfe, 2000; Baril, Ayman, & Palmiter, 1994; de Vries, Roe, & Taillieu, 1998; Farh, Podsakoff, & Cheng, 1987; Orpen & Hall, 1994; Podsakoff, Niehoff, MacKenzie, & Williams, 1993; Roskin & Margerison, 1983), but like Contingency Theory, it can be too complex for many organizations to put into practice in any useful way. There is no one theory of leadership or associated assessment instrument available today which can adapt itself to any leadership culture with an organization. So practitioners cannot simply define their particular culture or leadership challenges, find an assessment designed for their "culture type," and employ it.

Emotional Intelligence

Dan Goleman (1995) described Emotional Intelligence (EI), which offers up a whole new set of intelligence dimensions, by using a story in the introduction of his book:

> A friend was telling me about her divorce, a painful separation. Her husband had fallen in love with a younger woman at work, and suddenly announced he was leaving to live with the other woman. Months of bitter wrangling over house, money, and custody of the children followed. Now, some months later, she was saying that her independence was appealing to her, that she was happy to be on her own. "I just *do not* think about him anymore – I really *do not* care," she said. But as she said it, her eyes momentarily welled up with tears.

> That moment of teary eyes could pass unnoted. But the empathic understanding that someone's watering eyes means she is sad despite her words to the contrary is an act of comprehending just as surely as is distilling meaning from words on a printed page. One is an act of the emotional mind, the other of the rational mind. In a very real sense we have two minds, one that thinks and one that feels (p. 8).

Goleman asserted that Emotional Intelligence is a necessary competency for leaders. Goleman and his colleagues developed the Emotional Competence Inventory (ECI), which is designed to measure emotional competencies and positive social behaviors (Goleman, 1995). The ECI assesses 20 competencies that are organized into four clusters: self-awareness, social awareness, self-management, and social skills. The ECI employs 360-degree assessment techniques that can include self ratings, peer ratings, and supervisor ratings. Not enough research has been done on this relatively new construct to determine its long-term utility, but it does hold validity on its face, and is appealing to practitioners for that reason as evidenced by the litany of practitioner-oriented literature about it (Hughes, Patterson, & Terrell, 2005; Kravitz & Schubert, 2000; Lynn & Lynn, 2002; Mapes, 2000; Merlevede, Bridoux, & Vandamme, 2003; Ryback, 1998; Sala, Urch Druskat, & Mount, 2006). Does Emotional Intelligence as a construct, or the ECI assessment as a tool, offer practical guidance on leadership selection at this time? Scholars have not yet ruled out EI's efficacy, so it continues to be a potential predictor of good leadership going forward.

Summary of Leadership Theories

As one can easily see from this limited overview, there are simply too many leadership theories that contradict each other or are too complex for practical application in most organizations. As an unknown author once put it, "Theorists would sooner use each others'

toothbrushes than adopt each others' terminology." Yukl stated that "Most leadership theories emphasize one category more than the others as the primary basis for explaining effective leadership" (2002). For instance, Heifetz's (1994) selection of leaders to spotlight in his "Leadership Without Easy Answers" was driven by those leaders who not only met the needs of their followers, but who also elevated them. James O'Toole's (1996) selection of leaders to discuss in his book was driven by his concept of the Rushmoreans, a group of men who O'Toole believes represent a "school of values-based leadership dedicated to democratic change" (p. 21). A now famous quote from Bennis and Nanus (1985) sums up the efforts of researchers in defining leadership:

> Never have so many labored so long to say so little. Multiple interpretations of leadership exist, each providing a sliver of insight but each remaining an incomplete and wholly inadequate explanation. Most of these definitions *do not* agree with each other, and many of them would seem quite remote to the leaders whose skills are being dissected. Definitions reflect fads, fashions, political tides and academic trends. They *do not* always reflect reality and sometimes they just represent nonsense (p. 4).

Schön (1983) not only recognized the separation that exists between research and practice, but contended that the gap is hierarchical, with the research camp considering itself superior. If this is true, it probably does not sit well with most practitioners, and could be more reason for their resistance in adopting research results.

In the face of such disagreement within the scholarly world, limited time for reading, and the difficulty in reading and interpreting scholarly information, HR professionals and hiring managers have learned over the years to rely on their intuition to make selection decisions. As Schön (1983) said, "Many professionals become selectively inattentive to data that fall outside

their categories" (p. 43). Edgar Schein (1992), Nathan Glazer (1983), and Herbert Simon (March, Simon, & Guetzkow, 1993) have all addressed the gap between professional knowledge and the demands of practice.

This section provided an overview of predominant leadership theories. The next sections consist of an examination of implicit and explicit leadership theories in organizations and how they substantively feed an organization's decision-making processes regarding leadership selection.

Implicit and Explicit Leadership Theories in Organizations

As Northouse was settling on a good operational definition of leadership in 1997, Robert House (1997) and other theorists were complaining:

> For example, to this day, the dominant proportion of the more than 3,000 studies listed by Bass (1990) is primarily concerned with the relationship between leaders and their immediate followers, and largely ignores the kind of organization and culture in which leaders function, the relationships between leaders and superiors, external constituencies, peers, and the kind of product or service provided by the leader's organization (p. 409).

Edgar Schein is most often associated with discussions about organizational culture. In his book, *Organizational Culture and Leadership* (1992), Schein defined culture as:

> … a pattern of shared basic assumptions that the group learned as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems" (p. 12).

Schein asserted that culture and leadership are interrelated; if leaders do not operate with awareness about culture, the cultures will manage them.

The culture of an organization, business unit, and even an entire industry can be an important aspect to consider in the selection of leaders, and must somehow be built in the selection process in order to increase potential person:job fit. From where does culture derive? Can culture be measured, assessed, or predicted? If so, can these measures be compared to leader measures to ensure a better fit? Is it possible to examine one element within an organization or industry, such as hiring practices, and understand its culture? In his book, *The Web of Life*, Capra (1996) warned against the Cartesian paradigm of examining the parts to understand the whole. He explained that, in Quantum Theory, "we never end up with any 'things,' we always deal interconnections" (p. 30). Gestalt psychologists, he pointed out, "saw the existence of the irreducible wholes as a key aspect of perception" (p. 31). Additionally, Capra explains that the term "ecology" was derived from the Greek word "oikos," which roughly translates to the English word "household," with ecology being the "study of the Earth's household and the relationships that interlink all members of the Earth's household" (p. 32).

Schein (1999) cautioned against attempting to assess culture for reasons other than the interconnectedness of things. He believed the greatest risk in working with culture is oversimplifying it, asserting that any culture assessment must include a thorough examination of three elements: the organizational artifacts, espoused values, and shared tacit assumptions. "Most questionnaires that purport to assess culture deal with issues such as teamwork, superior-subordinate relationships, the degree of autonomy or empowerment employees feel, the level of innovation or creativity that they display" (p. 27). But culture, in Schein's estimation, cannot be measured in this way because the shared tacit assumptions people within the organization make are far too subtle and buried to lend themselves to a questionnaire. In fact, he believed that it is nearly impossible to expose shared tacit assumptions. Malcolm Gladwell (2005), in his attempt

to explain to readers that sometimes people know things without knowing how or why, discussed one approach to capturing underling assumptions through the use of an Implicit Assumptions Test (IAT). He summarized research which indicates that the IAT can demonstrate race and gender biases, and used that research to describe how car salesmen offer higher prices to female and minority customers than they offer to Caucasian males.

Accessibility to an IAT is probably far off for most companies. While some authors cite or purport to have instruments that can measure and characterize organizational culture, they seem as confusing and conflicting as leadership theories (Glisson & James, 2002; Goffee & Jones, 1998; Sridhar, Gudmundson, & Feinauer, 2004). If there is not a way to characterize organizational culture, then how can we measure or assess it? As this section has described, organizational culture is an extremely difficult construct to measure, even though most acknowledge its importance in leadership selection, at least at face value. The next section overviews ways in which researchers have attempted to understand culture and underlying organizational assumptions.

Inquiry into Culture

Schön (1983), in his examination of how professionals think in action, introduced the concept of “reflection in action” which includes inquiry, reframing, hypotheses, and testing. Although he did not provide great detail about how this approach might be applied, it has potential utility in exploring organizational assumptions, both stated and assumed. A researcher could ask questions about the culture (“the way people do things around here”), begin to frame and reframe a hypothesis based on patterns in the responses, then test the hypothesis. But this model does not provide a clear enough direction for researchers.

Grounded Theory

Kathy Charmaz, in discussing objectivist and constructivist grounded theory, stated, "The strategies of grounded theory include (a) simultaneous collection and analysis of data, (b) a two-step data coding process, (c) comparative methods, (d) memo writing aimed at the construction of conceptual analysis, (e) sampling to refine the researcher's emerging theoretical ideas, and (f) integration of the theoretical framework" (Denzin & Lincoln, 2000, pp. 511-512). She also stated, "Researchers can use grounded theory models with either quantitative or qualitative data, although these methods are typically associated with qualitative research" (p. 510). While this approach provides much more detail on implementation, some researchers might find the concurrent collection and analysis of data difficult.

Appreciative Inquiry

The concept of Appreciative Inquiry (AI) centers around the "power of the positive question" (Cooperrider, Sorensen Jr., Yaeger, & Whitney, 2001, p. 130). The authors describe it as "a social construction based on a sociorationalist paradigm as opposed to the paradigm of logical positivism" (p. ix). They reject the notion that good research involves solving problems, instead proposing a "second dimension" of accepting "multiple ways of knowing, each of them valid in its own realm when judged according to its own set of essential assumptions and purposes" (p. 86). There is a new trend of using AI in survey instrumentation; and the authors opined that this approach is particularly useful in assessing culture, given that AI "may transcend national cultural boundaries, may represent a common human experience, and may have potential for organizational change which is universal and not limited by national cultural values" (p. 138). In one potentially relevant case study within the book, researchers used

Appreciative Inquiry interviews, review of historical documents, observations, and surveys which they then thematically coded and through which they derived major themes.

Theory of Action

In his recent dissertation, Scott Allen (2006) modeled an exciting new approach to making organizational assumptions explicit based upon the work of Patton (1997). Allen cited Argyris and Schön (1978) as introducing the concept of *Theory of Action*, meaning that people not only espouse certain theories, but they also unknowingly employ discrepant theories of action, particularly when embarrassed or threatened. Patton described five steps for making implicit theories of action explicit. He contended that the researcher:

1. Makes the process of theory articulation understandable.
2. Helps participants be comfortable with the process intellectually and emotionally.
3. Provides direction for how to articulate espoused theories that participants believe undergird their actions.
4. Facilitates a commitment to test espoused theories in the awareness that actual theories-in-use, as they emerge, may be substantially different from espoused theories.
5. Keeps the focus on this to make the evaluation useful (p. 223).

This approach provides a great deal more understandable detail and on its face, appears more practical than other methods previously described, although it has not been employed in research aside from Allen's (2006) recent dissertation. It provides behavioral guidance to the researcher, and acknowledges that what people say and what they do could be different.

Each organization handles personnel selection differently, which is likely a function of the organizational culture. The previous section provided an overview of methods of inquiry into culture. The next section addresses the tools used in personnel selection with the assumption that culture plays a role in decision-making.

Personnel Selection

A major contributor to the dilemma of personnel selection is the matter of predictive criterion validity. In other words, does the selection tool used actually predict an individual's performance as a leader? Does an examination of the characteristics, traits, motivations, or past results of a job candidate help predict leadership success in the future? By exploring current selection criteria for leaders in the organization within the context of the way it understands and defines leadership success (which may systematically vary by organization, business unit, or even industry), practitioners can then begin to evaluate the predictive ability of their selection methods.

In Chapter One, the researcher offered an analogy between the hiring decision and a research question, detailing the various methods of inquiry that feed the decision-making process. This review of the existing literature as it pertains to selection methods and their predictive validity will provide more detail into those methods. Schmidt and Hunter (1998) conducted one of the most cited meta-analyses on personnel selection. They examined the validity of 19 selection procedures for predicting job performance used over the 85 years preceding their study. They contended that, "(a) the economic value of gains from improved hiring methods is typically quite large, (b) these gains are directly proportional to the size of the increase in validity when moving from the old to the new selection methods, and (c) no other characteristic of a personnel measure is as important as predictive validity" (p. 263). They found work sample tests (r=.54), structured interviews (r=.51), general mental ability tests (r=.51), and integrity tests (r=.41) to have the highest predictive validity of the 19 methods reviewed (p. 265).

In deciding what selection methods to employ, practitioners struggle not only with identifying those techniques which provide the greatest predictive validity, but also with legal

and practical issues. Fortifying the selection decision-making process with additional steps may provide more useful data, but it also adds to the time-to-hire statistic against which many HR practitioners are measured. Additionally, their selection methods can be challenged from multiple fronts, including by the Equal Employment Opportunity Commission (EEOC), the federal agency established by Title VII of the Civil Rights Act of 1964 and granted the authority to investigate unlawful employment practices and to seek civil and even criminal penalties. Under federal guidelines, selection practices may not have adverse impact against people in protected classifications such as sex, race, color, age, religion, national origin, pregnancy, disability or veteran status. Selection practices must be regularly validated to ensure their compliance with EEOC guidelines, a practice that can be confusing and time consuming. The Seventh US Circuit Court of Appeals' recent court ruling in Karraker v. Rent-A-Center Inc. removed the Minnesota Multiphasic Personality Indicator (MMPI) from the repertoire of organizations using personality testing in selection decision. The court ruled that this assessment, designed to diagnose mental impairments, violates the Americans with Disabilities Act which bars pre-hire medical examinations. Such legal proceedings can cost companies large amounts in attorney fees, employee back pay, fines, settlements, and more. Sometimes it is simply easier to rely on familiar "tried-and-true" methods of selection rather than to attempt something that is new and potentially risky.

Predictive Assessment Instruments

A 1999 survey report on workplace testing stated that 46 percent of respondent companies use psychological measurement, with financial services companies in the lead for testing job candidates ("American Management Association survey on workplace testing",

1999). So it is clear that practitioners are attempting to find ways to predict success in job candidates. What assessments are being used in the field for leadership selection, and what is their performance so far? In order to assess the efficacy of any instrument, a number of measures are employed. In general, an instrument should be found to be both reliable and valid.

Reliability

To be reliable, a test must lead "to the same or similar results, regardless of opportunities for variations to occur" (Nunnally & Bernstein, 1994, p. 214). The reliability of a test is usually evaluated in terms of its stability over time (test-retest) or the "extent to which each item is measuring the same variable" (Kline, 2000, p. 28) (internal consistency). To assess test-retest reliability, three common approaches (test-retest, parallel form, and split half) yield a correlation coefficient which provides insight into the degree of the relationship between the scores from each of the tests. The Pearson correlation coefficient ranges from "-1" to "+1" with the highest absolute value of "+1" indicating the strongest relationship possible. Internal consistency reliability is illustrated using the alpha coefficient, which shows the relationship among the items within the test and which is "interpreted as if it were a correlation" (Kline, 2000, p. 28).

Validity

An instrument's validity depicts the extent to which it measures what it is intended to measure. Validity types include face, content, criterion, and construct. Face validity assumes that the hiring manager and job candidate in an employment selection scenario both see the test items as acceptable and reasonable. Content validity, in an employment setting, relates to the degree to which the test items match the job specifications (Rust & Golombok, 2004). An instrument has criterion validity (including predictive or concurrent validity) when its variables can predict an outcome (Nunnally & Bernstein, 1994). Construct validity refers to whether an assessment

measures an unobservable construct that it purports to measure (Kline, 2000). Two important themes within the construct validity concept are convergent and discriminant validity. Essentially, a test item should correlate highly with measures to which it is similar (convergent); and it should not correlate with those items to which it is not similar (discriminant) (Rust & Golombok, 2004). The level of validity is usually expressed as a correlation, as described by Nunnelly (1994) regarding predictive validity:

> Correlations based upon a single predictor, save for some settings highly dominated by intelligence (general cognitive ability), rarely exceed .3 to .4 (a figure that is also typical of predicting academic success). People are far too complex to permit a highly accurate estimate of their proficiency in most performance-related situations from any practicable collection of test materials…Tests that have only modest correlations with their criteria (e.g. correlations of .30 and .40) can improve the average performance of personnel markedly under optimal circumstances, e.g., many applicants for relatively few positions. Of course, many mistakes will be made in prediction, but on the average, persons who score high on the test will perform considerably better than persons who score low on the test (pp. 99-100).

A test can also have internal or external validity. Internal validity refers to the extent to which confounding variables have been eliminated or reduced; tests with high external validity can be generalized to the population at large (Nunnally & Bernstein, 1994).

The Culture of Inquiry

As mentioned previously, one problem with measuring leadership potential and success is settling on one definition or theory of leadership. If it were possible to identify and agree upon

one applicable leadership theory, a method of useful measurement could possibly be derived. For instance, research related to the trait theory of leadership could be translated into practice by conducting tests, physical observation, or self-reporting. As focus has shifted more to the relations between leaders and followers, researchers have begun examining the behavior of and interactions among all team members. This could be measured by assigning researchers to observe and report on the team's activities, either freestyle or with the use of checklists (Seltzer & Bass, 1990, p. 54).

Research Methods

Seltzer (1990) provides a summary of the history and approaches up to 1990:

> By now, it is fair to say that every procedure known to social science in general has been applied specifically to the study of leadership. These procedures have included autobiographical analysis; biographical analysis; case studies; the evaluation of news records; memoranda, and minutes of meetings; the analysis of speeches; biodata analysis; studies of communication patterns; autologs and observers' logs of leaders' activities; ratings by observers, superiors, peers, subordinates, and clients; judgments of verbal protocols; and individual interviews. Increasingly, investigators are using two or more approaches to increase confidence in their efforts. (p. 55).

Researchers continue to study leadership in a large variety of ways, using the full repertoire of social science research methods. For instance, in 2002, Newcombe and Ashkanasy (2002) looked at how the congruence between facial expressions and verbal messages affects followers' perceptions of leaders using a seven-item measure. Practitioners are more likely to appreciate the study conducted by Koene *et al.* (2002) regarding how leadership style can impact organization's financial performance. This study examined 50 supermarkets in the Netherlands,

and showed a relationship between local leadership, financial performance, and organizational climate.

Researchers have evaluated various aspects of leadership including behavior, motivation, interactions, problem-solving ability, and even humor (Avolio, Howell, & Sosik, 1999; Decker & Rotundo, 1999; Grugulus, 2002). Many of the attempts to measure leadership have used assessments and questionnaires based on one particular theory, such as charismatic or transformational leadership (Bass, Avolio, & Goodheim, 1987; Fuller, Patterson, Hester, & Stringer, 1996; Knight & Holen, 1985; Mullen, Symons, Hu, & Salas, 1989; Seltzer & Bass, 1990; Shamir & Zakay, 1998).

Major Assessments in Use

A literature search on measuring and predicting leadership was recently conducted which resulted in a group of 188 available empirical studies spanning the period from 1954 to 2005. Sixty-six of these studies focused on examining the quality of new or existing instruments. For instance, Alban-Metcalfe *et al.* (2000b) studied the reliability and convergent validity of a new transformational leadership questionnaire using repertory grid technique and confirmatory principal component analysis. The other 122 articles generally assumed the validity of the assessment tools used in the course of exploring some aspect of leadership. As an example, Bliese *et al.* (2002) tested the hypothesis that leadership moderates the relationship between stressors and subordinate well-being, and used the Least Preferred Co-Worker Scale (LPC) as a measure of leadership.

Few of the studies in either group, however, took into consideration the organizational or industrial culture, the company's or industry's espoused or assumed leadership theories, or other work environment elements. Most often, researchers approached their studies with a certain

leadership theory they had adopted and set out to test it with subjects, or simply used an assessment instrument specifically devised with a particular leadership theory in mind apparently assuming it indicates a generally acceptable definition of leadership. However, Fiedler's Contingency Theory of Leadership suggests that certain leaders are required for certain situations (Fiedler, 1967). The concept that an experience, competency, motivation, or personality necessary to succeed in one leadership situation might be unnecessary or even destructive in another is highly relevant to leadership studies. But how does one assess this moving target?

Another problem with the studies is that only a small number of them have attempted to link the assessment instrument with actual indicators of effective leadership, such as achievement of goals, subordinate satisfaction, or supervisor ratings. Just as scholars have struggled with finding one good definition of leadership, researchers have grappled with identifying "good" leadership outcomes. Although there are limitations with many approaches taken in the studies, some of the more useful attempts involve supervisor ratings (Chemers, Watson, & May, 2000), goal achievement (Sala & Dwight, 2002), promotions (McClelland & Boyatzis, 1982), income (Judge & Cable, 2004), and subordinate satisfaction (Schriesheim, A, DeNisi, & S, 1981). Unfortunately, however, most of the instruments evaluated were not used in actual leadership situations. For instance, a number of studies were conducted in college environments in which students describe leadership, and not in genuine work environments in which the study subjects are exposed to variations in leadership behavior. This leads back to the concern of predictive criterion validity; if the studies were not conducted in an actual work environment, how can practitioners trust them to predict leadership success in the business world?

The following summarizes several of the most often cited instruments from the literature search.

Multi-Factor Leadership Questionnaire (MLQ)

The MLQ is a multi-rater instrument most often associated with Bass' Transactional and Transformational Leadership Theories (Stogdill, 1974). The MLQ is mentioned in 16 of the studies within the two study groupings from the literature search. The approach to evaluating the test involves administering the instrument to a number of subjects (n ranges from 52 to 3,786). The researchers then conducted statistical analysis (usually Confirmatory Factor Analysis) on the results to determine construct validity and internal reliability (The results were mixed, with partial support for the instrument.). In one study, the researchers compared the results of the MLQ to Yukl's Management Practices Survey (MPS) to determine their independence (It was somewhat confirmed.). In only two of the studies about the MLQ did the researchers attempt to assess predictive validity by comparing the test scores to leadership effectiveness data (subordinate intent to leave, organizational commitment) (Bycio, Hackett, & Allen, 1995; Tracey & Hinkin, 1998).

Least Preferred Co-Worker Scale (LPC)

The LPC is most often associated with Fiedler's Contingency Theory of Leadership (Fiedler, 1967). Empirical study results about the instrument have been mixed, with problems with internal consistency and construct validity heading the list of concerns (Ilgen & O'Brien, 1974; Kennedy, K, & Jr, 1982; J. K. Kennedy & Gallo, 1986; J. K. Kennedy, Houston, Korsgaard, & Gallo, 1987; C. A. Schriesheim, Tepper, & Tetrault, 1994; Shiflett & Samuel, 1981; Weissenberg & Gruenfeld, 1966). One study successfully verified the test-retest stability of the instrument (J. K. Kennedy & Gallo, 1986), but another study completed the following year

by one of the same authors cast doubt upon the validity of the LPC (J. K. Kennedy, Houston, Korsgaard, & Gallo, 1987). Not one of the research studies attempted to evaluate the predictive validity of the instrument as compared to external factors such as supervisor ratings or subordinate satisfaction. A newer, shorter version of the LPC has received good preliminary reports, but has not been tested enough to assert its efficacy at this time (Ashworth & Hazer, 1986).

Leadership Practices Inventory (LPI)

Posner and Kouzes (1993) created the LPI. The assessment is in its third edition, and the authors claim it is the best-selling and most trusted leadership instrument available on the market today ("LPI Online - A Leadership Challenge Resource", 2002). It is a 360-degree instrument, which helps to address rater bias concerns. Researchers have evaluated this instrument using factor analysis, usually confirmatory, and have found some level of construct and criterion-related validity (Carless, 2001; Fields & Herold, 1997; Lam & K, 1998; Posner & Brodsky, 1992; Tourangeau & McGilton, 2004). However, the criterion-related validity claim seems somewhat unjustified in light of the fact that the test results do not appear to have been compared to any business-oriented measures of successful leadership. For instance, Tourangeau *et al.* (2004) conducted exploratory principal component analysis, regression analysis, and exploratory factor analysis using the assessment in relationship to leadership *theory*, not actual predictors of successful leadership. There is also a student version of this instrument (Posner, 2004; Posner & Brodsky, 1992). Using factor analysis, the researchers found it to be effective in differentiating between more and less effective student leaders, although they do not discuss the measurements used for leadership effectiveness to any useful extent.

Substitutes for Leadership Questionnaire (SLQ)

This assessment is associated with Kerr and Jermier's Substitutes for Leadership Theory (Kerr, 1978). The empirical findings regarding this theory and the research conducted so far using the SLQ instrument have been disappointing (Podsakoff, MacKenzie, & Bommer, 1996; Podsakoff, Niehoff, MacKenzie, & Williams, 1993). Some studies found a reasonable amount of construct validity and internal reliability in the SLQ or its revised version (Houghton & Neck, 2002; Podsakoff & MacKenzie, 1994) and another study's results supported 11 of Kerr's 13 factors (Pitner, 1988), while yet another identified issues with the construct validity, stating that 10 of the 13 subscales are not reliable for research use (Williams, Podsakoff, Todor, & Huber, 1988).

Leader Behavior Description Questionnaire (LBDQ)

The LBDQ was created by the Personnel Board at Ohio State University, and is often called "The Ohio State Assessment." The instrument allows team members to describe the leader's behavior. Ohio State University provides certain versions of this instrument free of charge, so it can be a very affordable option for researchers and practitioners. The research conducted on this instrument has generated mixed results, with some research finding limited evidence of construct validity (Follert, 1983), other researchers asserting that the "96-item questionnaire could be used with confidence" (Gioia & Sims, 1985; Spangenberg & Theron, 2002), and yet another study expressing concern about the instrument, having found it to be susceptible to contamination by social desirability or through a priming effect (Head, 1991; Tracy, 1987). Edwards (1957) described social desirability as the tendency of subjects to "fake good" or "fake bad" in questionnaires, depending upon the desired outcome. "Priming" is a psychological term used to describe the phenomenon in which an event someone has

experienced facilitates (or, in the negative case, impairs) his or her processing of a subsequent event (Tracy, 1987).

Leader Effectiveness and Adaptability Description (LEAD)

The LEAD is a relatively new instrument, usually associated with Kerr and Jermier's Substitutes for Leadership Theory (Kerr, 1978). Researchers found serious test-retest reliability and construct validity problems with this instrument, and offered suggestions for making the assessment more efficacious (Butler, 1993; J. E. Edwards, Rode, & Ayman, 1989; Lueder, 1985). One of the reasons the LEAD may be difficult to validate is that it is not simply measuring leadership behaviors, but also those variables which affect the leader-subordinate relationship and work environment. These types of factors are much more difficult to ascertain, yet this assessment will be very important if the problems can be resolved over time. The fact that the LEAD focuses on more than just the leader's traits (relationships and context) is a step in the right direction for assessment instruments that practitioners will see as useful. However, because every work environment and supervisor:employee relationship is somewhat different, the LEAD is still not a complete solution because, like most other available assessments, it is not customizable to a variety of work environments or relationships.

Transformational Leadership Questionnaire (TLQ)

The TLQ is a newer instrument, most frequently associated with Bass's Transformational Theory of Leadership (Stogdill, 1974). The TLQ was preceded by the MLQ, discussed earlier in this document, and has mostly been tested in the United Kingdom. Researchers assert that this assessment is more valid and reliable than the MLQ (including convergent and discriminant validity and criterion-related validity) and is easier to administer and take (Alban-Metcalfe & Alimo-Metcalfe, 2000a, , 2000b; Alimo-Metcalfe & Alban-Metcalfe, 2001). One potential

problem is that, while different aspects of transformational leadership were significantly correlated with each of the five criterion variables, researchers could not replicate this for different groups of managers (Alban-Metcalfe & Alimo-Metcalfe, 2000b). This problem supports concerns about the ability of any one assessment to accurately predict the success of leaders in different organizations.

Summary of the Assessments

As one can see, the results of research assessing the efficacy of each of these tests have been mixed. It appears that no one assessment is a panacea for assessing leadership. This could be attributed to the fact that most assessments have not been adequately compared to measures of successful leadership within the organizational context, and because, of all the factors which contribute to the success of a leader, one of the most important is the context or work environment in which the leader must operate. The field of leadership study stands to benefit from more research on assessments compared with more objective data about successful leadership outcomes in actual work environments.

The Case Against Assessments

Critics of assessment instruments offer a broad range of objections. For instance, Paul's (2004) skeptical book entitled *The Cult of Personality: How Personality Tests Are Leading Us to Miseducation Our Children, Mismanage Our Companies, and Misunderstand Ourselves,* sets forth her concerns. She was disquieted by the now-outlawed MMPI's items about sexual habits, bowel movements, and other non-work-related topics; the fact that higher level managers are often not tested yet they are the root of most corporate issues; gender, racial, and other biases; the lack of deception scales in most assessments; and the overly broad nature of the Big Five personality tests. Perhaps her greatest concern; however, was the irrevocable harm that people

endure when improperly classified. “Measurement is a useful tool,” she quoted from Gordon Allport, “but if it makes one think one has embraced the totality of a personality by having a series of scores, then it has gone too far” (p. 214).

Perceptions of the Test-Taker

Paul is not the only person who is apprehensive about personality assessments. Harland (2003), Folger and Cropanzano (1998), and Gilliland (1993) have all helped to raise awareness about the perceived fairness of selection systems and assessments from the standpoint of the job candidate. When an applicant perceives that he has been treated unfairly, he is more likely to decline a job offer and to complain to other potential job applicants about the process. Flippant instructions, such as “*do not* think too long about any one question” can give the test-taker the impression that the test-giver will be the one to do the thinking (Braun, Jackson, Wiley, & Messick, 2002). Gilliland (1993) offered suggestions for improving the test-taking experience for candidates, including explaining the purpose of the test in advance, providing results along with a thorough explanation of their meaning, and eliminating the repetitive questions that some test takers find offensive (Braun, Jackson, Wiley, & Messick, 2002).

User Error

Even if the assessment instrument itself is perfectly viable, misuse, misinterpretation, or bias on the part of a hiring manager can lead to poor results including increased employee turnover, hiring costs, and even law suits (Heneman III, Hamstra, & Brown, 1980). Furthermore, some personality dimensions might lend themselves to assessment more readily than others. Viswesvaran *et al.* (1996) cited research suggesting that leadership and communication competence are more difficult to evaluate than dimensions such as output and errors, and offer examples in which raters evaluated ratees much more accurately on some dimensions than

others. These concerns can be addressed through proper training on the appropriate uses and scopes of assessments.

Type I and Type II Selection Errors

Other problems perplex those who might use assessments in selection decisions. Heneman III *et al.* (1980) have written about Type I and Type II errors as related to assessments. Just as scholars face this issue in statistical analysis, hiring managers risk using assessments which might yield a false negative (the candidate failed the test but would have been a good fit) or a false positive (the candidate passed the test but is not a good fit), respectively. Imagine the frustration a manager might experience when trusting an assessment's results yields a candidate that is not at all suitable. Just as in statistical analysis, attempts can be made to mediate the risk of Type I and Type II errors. But, due to the fact that these errors are inversely proportional, the challenge is in striking the correct balance (Jaeger, 1993).

Test Bias

The term, "disparate (adverse) impact" is used to describe the outcome of a selection process that appears to be nondiscriminatory, but which excludes certain classifications of people disproportionately (Nunnally & Bernstein, 1994). A number of assessments in the past have had this result, including earlier versions of the Stanford-Binet Intelligence Quotient instrument (Braun, Jackson, Wiley, & Messick, 2002). Tests can also be biased through use of language including local colloquialisms, references to monetary amounts in another country's currency, and even by using examples and questions which would tend to resonate more with one population than another. The Equal Employee Opportunity Commission has addressed and enforced many such issues as discussed elsewhere in this document; guidelines are available to employers in order to reduce the risk of illegal discrimination in employment decisions.

"Faking" the Test

In assessments with overt items, it is easier for test takers to consciously or unconsciously present themselves in a favorable way because the desired answer to the question is obvious. While some assessments such as the MMPI have a built-in "lie detector" or distortion scale, others such as the Fundamental Interpersonal Relations Orientation – Behavior (FIRO-B) do not. Even when a validity or honesty scale is available, clear instructions are normally not provided to practitioners regarding how to interpret the scales and how they affect the validity of the assessment results. Test developers address this issue through the development of more covert questions and the use of distortion scales; however these techniques are not always sophisticated or covert enough to outsmart a savvy test-taker, or to be clear to those practitioners responsible for interpreting assessment results.

Halo Effect

Nunnally and Bernstein (1994) discussed one potential problem with behavioral ratings, "…a rater may confound the specific attribute to be rated with other attributes, including an overall evaluation, producing a halo effect" (p. 339). The term "halo effect" has been described as "a rater's tendency to perceive an individual who is high (or low) in one area as high (or low) in other areas as well (Wells, 1907; Thordike, 1920). It reflects a failure to discriminate among conceptually distinct and potentially independent aspects of a ratee's behavior (Sall, Downey, & Lahey, 1980, p. 450) and is a form ofvariance-induced reduction in the divergent validity of ratings…" (p. 373). The halo effect can make it difficult for researchers to single out specific predictors of behavior; however, some statistical approaches to measuring, evaluating, and even correcting for the halo effect are available (Oh & Ramaprasad, 2003).

Rater Bias

The matter of rater bias becomes an important issue as researchers attempt to find expedient ways to predict leadership potential.

Self-Rating

Certainly, one of the easiest ways to assess leadership potential is to have the job candidate herself complete the assessment. However, several realities can make this problematic. First, social desirability can play a large role in distorting assessment results (Braun, Jackson, Wiley, & Messick, 2002). "Although its role is clear in determining overall differences in response to items, its status as an individual difference variable is somewhat debatable. Moreover, it is important to separate the tendency to give a socially desired response with lack of self-knowledge" (Nunnally & Bernstein, 1994, p. 340). Test takers may consciously or unconsciously provide responses that they believe would cause the entity requiring the test to hold them in higher regard. Or they may simply not be sufficiently self aware to accurately rate themselves on certain personality or other dimensions. Braun *et al.* (2002) suggested that a self rater's denial of a behavior or trait could have one of two meanings: the absence of the trait, or the very opposite. In fact, they state that self description in and of itself is paradoxical. For instance, agreeing to the statement "I am modest" could actually demonstrate immodesty – the self rater can find herself in a double bind in these instances.

Supervisor Ratings

Sala (2006) studied executive performance as evaluated by the executives themselves and other raters as well. The ratings from the executives' supervisors and direct reports were found to be the most strongly related to actual performance, while self ratings and peer ratings were not as reliable. In their meta-analysis on this topic, Viswesveran *et al.* (1996) noted that for supervisory

ratings, overall job performance was most reliably rated (r=.86). However, Tziner *et al.* (2005) produced evidence that supervisory "rating accuracy has more to do with the deliberate, volitional distortion of performance ratings than was previously recognized…" (p. 89), attributing the distortion to a combination of uneasiness with performance appraisal processes and the need for supervisors to achieve their own personal goals. The reliability of supervisor ratings can be improved when the supervisors rate or rank their direct reports *in relation to* other employees and when supervisors provide narrative detail about subordinates' performance in addition to a numerical rating (Viswesvaran, Ones, & Schmidt, 1996).

Range Restriction

MacCann *et al.* (2003) noted problems that arise in scoring assessments and questionnaires. "Even if the standard error is quite small, a relatively large proportion of people cluster on a roughly equal score (plus or minus one standard error for example)" (p. 253). They suggested that weighted algorithms might be used to address skew and kurtosis to correct for range restriction. The issue of range restriction can be extremely problematic for researchers and practitioners alike. For instance, if a performance assessment includes a scale of one (worst) to five (best), supervisors will tend rate their reports in the "3" category (central tendency), which makes it more difficult to discriminate between employees. "Questions which everybody answers in the same way obviously cannot cast light upon how people differ, regardless of whether the content of the item makes it too easy or too difficult on an abilities test or the wording leads everyone in a common direction on any type of test" (Nunnally & Bernstein, 1994, p. 377).

Addressing the Concerns

This dissertation addresses some of the key concerns regarding the use of assessments including rater bias, test bias, and selection error. In spite of these and other apprehensions in using assessment instruments, the potential benefits of employing them successfully far outweigh the negatives. So long as researchers and practitioners are aware of the potential pitfalls and take steps to address them, most of the risks and problems can be effectively mitigated.

Leadership Assessment in the Healthcare Industry

Sixteen percent of the gross domestic product in the United States was earmarked for healthcare in 2004 (C. Smith, Cowan, Heffler, & Catlin, 2006). Healthcare costs continue to rise, as do bureaucratic, insurance, and legislative pressures for those in the field. Leaders of healthcare organizations find themselves in a unique position because they are not only accountable to their organizations for their strategic decisions, but also to the community and society at large (Schultz, 2004). Because healthcare organizations are "highly professionalized and require highly personalized interactions" (Irvine, Leatt, Evans, & Baker, 1999, p. 79), the demand for specialized assessments that are customized for the healthcare industry can be strong. Additionally, the required competencies, skills sets, and behaviors for successful leadership in a healthcare environment can differ greatly from those of a manufacturing, corporate, or other industry. Because "Physicians and nurses have traditionally been encouraged to act in a self-directed fashion" (Irvine, Leatt, Evans, & Baker, 1999, p. 80), the required style of management tends to be different from that required in other industries. Furthermore, the advent of managed care, outcome assessment, and alliances between healthcare organizations has created a quickly changing and highly competitive environment which "…makes strategic management of healthcare organizations even more challenging, and the changes are expected to continue in the

near future" (Schultz, 2004, p. 104). Schultz (2004) notes, "In the healthcare management literature, however, we find that little attention is paid to the quality and impact of strategic decisions made by top executives" (p. 104).

A recent search yielded literature describing the use of both established and newly created leadership assessments in the healthcare industry. Several new assessments were validated in the field, measuring constructs including empowerment (Irvine, Leatt, Evans, & Baker, 1999; Klakovich, 1995), 360-degree feedback (Garman, Tyler, Darnall, & Lerner, 2004), the quality of interaction between leaders and followers (Bhal & Ansari, 1996), and supportive leadership (McGilton, 2003) in nursing and other healthcare environments. Several studies reported field investigation of existing instruments such as the MLQ (Tejeda, Scandura, & Pillai, 2001; Vandenberghe, Stordeur, & D'hoore, 2002), the LPI (Tourangeau & McGilton, 2004), and the assessment developed by Podsakoff *et al.* (1984) to measure leader reward and punishment behaviors. The number of newly created instruments being validated in healthcare organizations is almost twice the number of existing, non-healthcare-oriented assessments cited in the literature review.

Among the newly developed instruments is the Healthcare Leadership Inventory (HLI). Developed by The Kingwood Group, the HLI was first used in late 2005. As its title implies, the HLI is intended to address the special concerns that HR practitioners in the healthcare industry face when trying to hire and promote good leaders. Unlike the other assessments described in this chapter, the HLI was created exclusively for leadership selection and development in the healthcare field. Based in part upon the Big Five Personality Theory, the instrument also includes critical thinking and customer service orientation dimensions. The assessment is so new that predictive validity has not yet been assessed; however, it has shown a promising Cronbach's

Alpha (internal reliability consistency) in the range of r=.73. The Kingwood Group conducted a validity test comparing its scores to supervisory ratings of performance. The results are encouraging with correlations between HLI factors and performance as high as r=.30 for self-confidence and overall performance (J. E. Smith, 2006). While this assessment is still very new, it holds some promise for practitioner use in the healthcare field.

Given the unique nature of the healthcare work environment; healthcare professionals who are promoted to leaders (e.g. nurses and physicians with little or no traditional management or finance experience); and the changing industry, it is not difficult to understand why hiring managers and human resources practitioners might want (and need) assessment instruments whose results are more likely to predict leadership success in their organizations. For an instrument to be truly applicable to the healthcare industry, not only must the test items themselves be oriented to the medical profession, but the results must also be "normed" *vis-à-vis* other healthcare leaders as opposed to leaders of other industries. Normalizing (or norming) is the process through which large-scale assessment results are standardized within a specific job group, company, region, or industry (Nunnally & Bernstein, 1994).

Leadership and Promotion Potential

What exactly is leadership potential? Rogers and Smith (2004) asserted that there are four cornerstones of leadership potential: Leadership Promise (the propensity to lead others), Personal Development Orientation (including receptivity to feedback), Mastery of Complexity (including adaptability and conceptual thinking), and Balance of Values and Results (this factor is unique to organizational culture) (Rogers & Smith, 2004). Higgs and Aitken (2003) found a relationship between emotional intelligence and leadership potential. Just as the variety of leadership theories and organizational cultures makes it difficult to define leadership, it is similarly difficult to

define leadership potential or to develop a description of “effective” leadership. A proposed operational definition is simply that *leadership potential is the antecedent to success as a leader in a particular organization or industry*. Because leadership success is defined differently depending upon organization and industrial culture in addition to other considerations, ultimate identifiers of leadership potential, if they exist, cannot be found without examining a particular organization or industry. In fact, research shows that people are more likely to rise to their potential when their supervisors believe in and hold high standards for them (Eden & Shani, 1982; Rosenthal & Jacobson, 1968; Sutton & Woodman, 1989). So if an employee demonstrates those traits, skills, or behaviors that his supervisor believes to be indicative of leadership potential (based on the implicit leadership theories of the supervisor and the organization/industry), the supervisor is more likely to provide future leadership opportunities to that employee and see that employee as having strong potential for promotion. To demonstrate this point for the more visual reader, Figure 2 is provided.

Figure 2.

Interplay of Leadership Theory and Leadership Opportunities

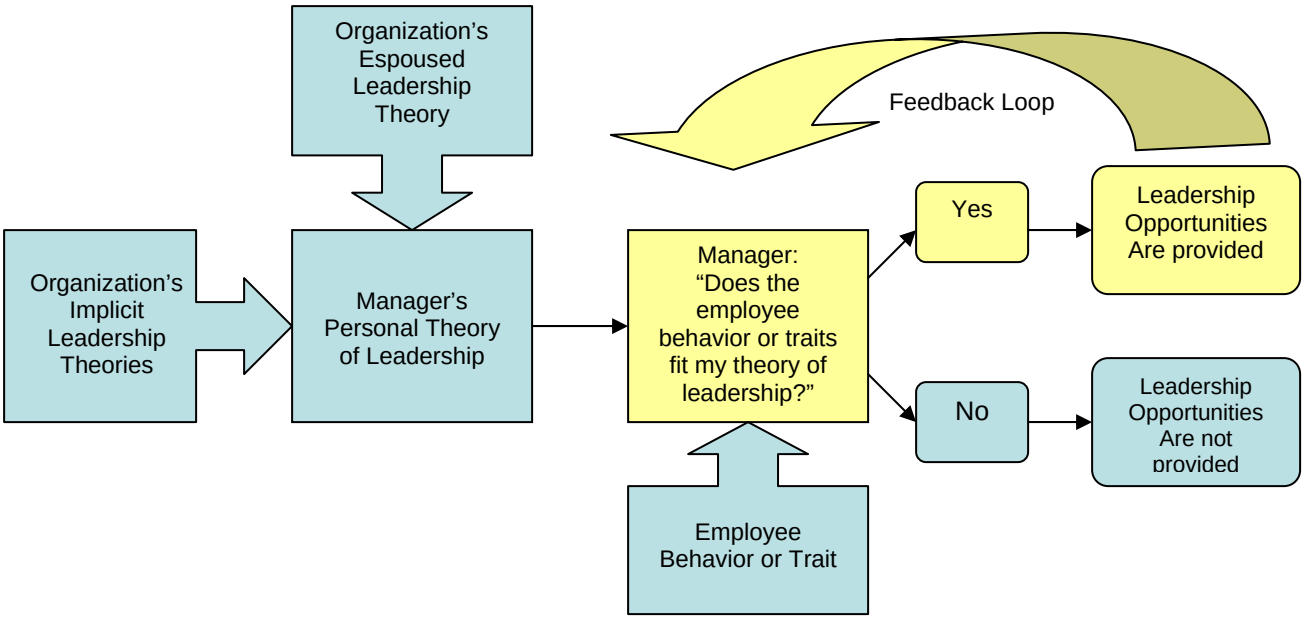


Figure 2 visually describes the influence of the organization's espoused and implicit leadership theories on a manager's personal theory of leadership, and the interplay between the manager's paradigm of leadership, subordinate behavior, and leadership opportunities. Those employees who demonstrate those behaviors that the manager associates with leadership (whether accurate or not) tend to have more leadership opportunities. This leads to a circular process in which those who have more opportunities for leadership improve their skills and advance into leadership roles. Based upon this logic, an employee's potential for promotion could be used as proxy measure for leadership effectiveness.

Summary

Although practitioners expend increasing resources to identify good potential leaders, selection errors continue to occur and can be quite costly. There are so many different, complex,

and conflicting theories of leadership that practitioners have difficulty subscribing to any one. Matching a leader to the environment is another challenge for practitioners, particularly within the healthcare industry. Current research can inform organizational selection practices, but most practitioners do not apply the results of academic research within their organizations. To the extent that researchers or practitioners can identify predictors of leadership for the healthcare industry, they can bridge the gap between scholars and practitioners. Practitioners can then apply these lessons learned to reduce selection errors.

Chapter Two reviewed the literature on selection methods, personnel selection, and the use of assessments. In Chapter Three, the methodology that was employed in this research will be discussed, as well as support for why the chosen design was appropriate for the stated research questions.

CHAPTER III - METHODOLOGY

Introduction

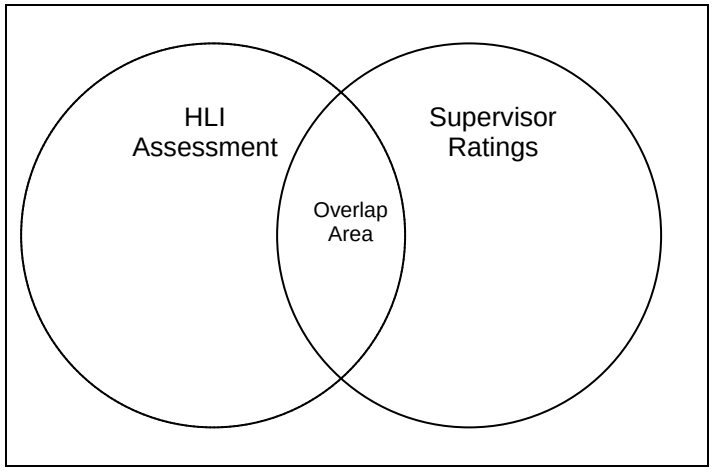
Chapter Two provided a review of the literature and an overview of applicable theories regarding leadership selection. The purpose of this chapter is to identify how the researcher planned to investigate the degree of congruence between leaders' performance ratings of their subordinate subjects versus the subjects' scores on the Healthcare Leadership Inventory (HLI) assessment. The overall goal of this analysis was to identify predictors of leadership potential using data provided by the publishers of the HLI assessment that were used in the original validation study in 2005.

Research Design

The study utilized descriptive quantitative research to examine correlations and relationships among its variables. Specifically, ratings provided by immediate supervisors were compared to HLI assessment scores. The researcher also conducted statistical analysis of the performance evaluation scores to explore implicit leadership theories that the immediate supervisors may share. The intended purpose was to identify differences between the high- and low-potential managers and to better inform leadership selection practices in the future. A conceptual map is provided in Figure 3.

Figure 3.

Conceptual Map of the Research



As Figure 3 shows, the "Overlap Area" between the HLI Assessment and the Supervisor Ratings is hypothesized to consist of the predictors of leadership potential. This overlap is the primary area of focus for the current research.

The Research Questions

This research study sought to answer the following questions:

RQ1: What factors within the performance evaluation instrument are significantly related to promotion potential as rated by the subjects' supervisors?

RQ2: What factors within the HLI instrument are significantly related to promotion potential as rated by the subjects' supervisors?

RQ3: How do the performance evaluation predictors compare to the HLI instrument predictors?

RQ4: What other elements of the assessment instrument are significantly related to leadership performance ratings?

Population

The Internet-based version of the HLI assessment was administered to a convenience sample of 375 managers who worked at 16 hospitals in the United States which agreed to participate in the research project. Participation of the management employees within each hospital was voluntary. The sample was reduced to 195 employees from 11 hospitals due to the lack of supervisory ratings. The managers' organizational levels varied from front-line supervisors to senior management within both clinical and non-clinical hospital settings. Participating hospitals ranged in size from 76 to 419 beds, with an average bed size of 178. Three of the hospitals were located in Ohio; two were in Michigan and Kansas; and the others were in Louisiana, West Virginia, Illinois, and North Carolina.

Ethical Issues

As all of the data used in this research are archival, this study did not pose any additional benefits or risks for the managers who were assessed and rated. Their immediate supervisors had already assessed their performance, and presumably had already defined programs for addressing managers with low and high leadership potential.

The assessment and performance evaluation data for managers were held confidential and only used in the manipulation of the assessment data. Any confidential files or documents, printed or electronic, were destroyed once the research was complete. A summary of findings was provided to The Kingwood Group.

Instrumentation

Healthcare Leadership Inventory

The HLI is a globally available assessment created by The Kingwood Group specifically for the healthcare industry. The assessment development was based upon literature on leadership

selection, existing databanks of The Kingwood Group's assessment results, and client input. The Kingwood Group has collected considerable evidence that the instrument can assess 10 different work-related performance factors including Critical Thinking, Achievement Orientation, Conscientiousness, Customer Orientation, Emotional Evenness, Innovative Thinking, Multi-Tasking, Openness to Change, Self Development, and Self Confidence. The most common applications are assessments for selection and development of managers and supervisors within the healthcare industry, particularly within a hospital setting. The assessment includes constructs consistent with Big Five Personality Theory, including Extroversion, Neuroticism, Agreeableness, Conscientiousness, and Openness factors. In the early development stages, the instrument had more than 20 factors which were pared down through an iterative process of validity checking to 10 factors. The HLI instrument can be taken in 60 to 90 minutes, and consists of 183 items including statements such as "It is easy for me to re-prioritize my work to meet changing needs" and "I win most arguments." Subjects rated each item, many of which are customized to the healthcare industry, on a five-point Likert scale from "Strongly Disagree" to "Strongly Agree." The Critical Thinking factor derives from a stand-alone instrument whose items have been customized to the healthcare industry. For this category, there is an additional 68-item instrument that has been construct-validated with the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1942) with a correlation of r=.55.

Reliability and Validity

Reliability and validity tests for the HLI assessment were conducted in January 2006. In its initial development stages, the instrument was administered to 68 employees who were predominantly from the healthcare industry, then administered a second time approximately three months later. The average Pearson product-moment correlation coefficient was r=.73

($p<.05$). The construct validity of the HLI assessment was evaluated *vis-à-vis* the well-established Neuroticism-Extroversion-Openness (NEO) Personality Inventory (based upon the Big Five Personality traits) and showed correlations ranging from r=.43 for Openness to Change to r=.77 for Achievement Orientation (J. E. Smith, 2006).

After the initial development phase, The Kingwood Group began the second phase by collecting the data which were subsequently provided for the current dissertation. In these tests, the HLI had a Cronbach's Alpha scale for internal consistency reliability ranging from .71 for Conscientiousness and Openness to Change to .87 for Innovative Thinking (J. E. Smith, 2006). Concurrent validity was assessed using performance evaluations completed by the subjects' immediate supervisors, with Pearson correlation coefficient scores ranging from r=.15 to r=.30 ($p<.05$), and with Cronbach's Alpha scores ranging from .71 to .87 for the 10 factors (J. E. Smith, 2006). Because the HLI is so new, predictive validity tests have not yet been completed. The technical manual for the instrument is in press (J. E. Smith, 2006). The assessment's items are proprietary to its creator, but a sample HLI Selection Report is included in Appendix A. The first two pages of the selection report provide scored results of the HLI assessment; the subsequent 13 pages provide structured interview tools including custom questions for which to probe areas of concern from the assessment.

Adverse Impact

Human resources practitioners may be familiar with measures of adverse impact as provided by the Equal Employment Opportunity Commission in the form of *Uniform Guidelines on Employee Selection Procedures ("Code of Federal Regulations, 41 CFR 60 - 3.4 - Information on impact", 2006).* According to the guidelines, employers may not use selection methods which have adverse impact on job candidates within federally protected groupings, such

as race, age, and gender. The Kingwood Group has conducted the federal government's recommended "4/5th Rule" evaluation of adverse impact and has found no adverse impact by race, age, or gender for its assessment (J. E. Smith, 2006).

Performance Evaluations

The immediate supervisors of the test takers completed a 12-item performance evaluation about the test takers. Ten of the 11 items were set to a seven-point Likert scale, and included Drive for Results, Conscientiousness, Customer-Orientation, Emotional Evenness, Innovative, Multi-tasking, Openness to Change, Self Development, Self Confidence, Potential for Promotion, and Overall Job Performance. The eleventh item was a ranking score in relation to other managers, including the options: "In the top 5%," "In the top 10%," "In the top 25%," "In the top 50%," "In the top 70%," and "In the bottom 30%." All items on the performance evaluation were related to the factors assessed in the HLI (having the same or similar name), with the exception of Potential for Promotion, Overall Job Performance, and Ranking. The performance factors were not weighted in any way.

The immediate supervisors received no special training in how to complete the performance evaluations, nor did they have information on the HLI assessment scores of their direct reports. The performance evaluation document provides a one-sentence description of each item, such as the descriptor for "Potential for Promotion: Capable of being promoted to the next level of management." The degree to which any of the immediate supervisors has or has not had formal training on performance ratings and/or rating errors, and the degree to which the subject managers have taken previous assessments unrelated to the HLI cannot be estimated because they worked at a number of different hospitals with presumably varying levels of management training or awareness on such content. So, to some extent, the performance evaluation form

completed by the immediate supervisor captures a level of intuition, particularly relating to the Potential for Promotion item. The importance of this subtle fact will become apparent to the reader later in this chapter when the data analysis is discussed.

It is also essential to note that the performance evaluation form used was not part of any existing company systems at the participating hospitals; it was created for the purposes of evaluating the validity of the HLI instrument. The performance evaluation form was not used by the participating organizations to justify pay changes, promotions, or even disciplinary actions. The supervisors were not required to review the evaluations with their subordinates or provide any explanation to them about the process whatsoever. Murphy and Cleveland (1995) produced evidence that "ratings collected for administrative purposes (for example, salary administration) are significantly higher than ratings of the same individual collected for other purposes, such as feedback or research" (p. 246). Because "linking appraisals to personnel decisions, such as compensation and promotions, leads to higher incidence of distortion and rater errors" (Coens & Jenkins, 2000, p. 26), the performance data used in this study may be less contaminated than an artifact of an internal process might be. The Performance Evaluation form is included in Appendix B.

Instrumentation Summary

The researcher's decision to use the HLI Assessment in this research is based upon an examination of other instruments used for assessing leadership inside and outside of the healthcare industry, its breadth and depth, the fact that the critical thinking portion is customized for a healthcare setting, its encouraging initial validity and reliability studies, the development report that is provided to the test-taker, and its straightforward description of the "cheat scale." Furthermore, the assessment's relationship to the Big Five Personality Theory and its cultural

universality provide the potential for it to be used internationally (John & Srivastava, 1999), which can be very important in today's global economy. It is the researcher's hope that an outcome of this research is an increased understanding of the HLI instrument and its uses in predicting leadership potential.

Data Collection Procedures

The data used for the study derived from two archival sources: performance evaluation data from the immediate supervisors of the test takers, and the HLI assessment instrument scores. Data collection was conducted by TestSource, an affiliate of The Kingwood Group, for the purposes of conducting reliability and validity tests on the instrument. The management of TestSource approached the hospitals that currently use the Healthcare Selection Inventory (HSI), also created by The Kingwood Group, using a solicitation letter. The solicitation letter was augmented with discussions about the new assessment during site visits at some hospitals.

For the supervisory rating data, TestSource emailed the performance evaluation form to its contact at each hospital. The hospital contacts distributed an evaluation form to each rating supervisor, along with a facsimile cover sheet which the supervisors then used to return the evaluation data. A TestSource employee then manually entered the supervisory ratings into a spreadsheet file, and sent it to The Kingwood Group for analysis. Instructions and a link for taking the web-based HLI and Critical Thinking assessments were also provided to the participating hospitals via email, and were then provided to the subject managers. The subjects accessed the HLI and Critical Thinking instruments online via the TestSource web site. An employee from TestSource exported the test data from the web-based interface to a Microsoft Excel spreadsheet which was sent to The Kingwood Group for analysis. Data collection began in the spring of 2005, and completed in October 2005.

Performance Evaluations

As discussed in the instrumentation section, the immediate supervisors of the test takers were asked to complete a performance evaluation form for each of their direct reports who held management positions. Supervisors who provided performance information did not have access to assessment scores, eliminating the possibility of criterion contamination. The Kingwood Group provided the results by manager - separate performance ratings for each item - in the form of an SPSS file (the same file that included the HLI scores) to the researcher for data analysis.

Healthcare Leadership Inventory

The HLI instrument was administered to 375 managers at 16 different hospitals in the healthcare industry during its initial validation studies in 2005. In addition to the assessment items, each subject was invited to provide demographic information such as gender, race, age (over or under 40 years old), time in position, time at hospital, and management level (front-line supervisor to vice president). The Kingwood Group provided the assessment results in the form of an SPSS file which listed all subject managers by name with their supervisors' ratings for each of the factors. Although the instrument underwent a validation study in 2005, data analysis described here takes a different direction and goes deeper than that conducted in the original study. Specifically, while the original validation study did include a correlation analysis of the HLI factors and certain of the supervisor ratings including Promotion Potential, multiple regression was not conducted to identify a potential cluster of the best predictors. Additionally, the chief analytical focus was using a "Computer Performance" score which was a combination of Promotion Potential, Overall Performance, and Rank. Although this focus was useful for the initial validation study, this dissertation focuses on the Promotion Potential facet.

Data Analysis

Following access to the database, the data were analyzed using the SPSS statistical package (Grad Pack Windows Version 14), which has been used in similar research (Stone, 1993). Presentation of the results was achieved using descriptive statistics, such as means, sample numbers, and standard deviations. All statistical analyses were conducted using an alpha level of $p<.05$. Factorial ANOVA was also conducted to evaluate differences in scores by gender (male or female), age (under 40 or 40+), and ethnicity (Caucasian or non-Caucasian) and to identify any main or interaction effects.

The first step was to answer ***Research Question One: What factors within the performance evaluation instrument are significantly related to promotion potential as rated by the subjects' supervisors?*** In other words, what factors appear to be linked to promotional potential based on the immediate supervisor ratings? First, a check was conducted for any redundant factors through bi-variate correlation analysis. This step is intended to identify and separate out the influence of variables which duplicate the predictive ability of other variables (George & Mallery, 2006). Next, using the potential for promotion scores as the dependent variable, stepwise multiple regression analysis was conducted using the other nine supervisor ratings and the ranking score as possible explanatory variables to determine the extent to which they co-vary with the promotional potential factor. Stepwise regression is the method of choice for explaining the combination (or model) of factors that significantly contributes to the prediction of the dependent variable. This process excludes those independent variables that do not provide additional predictive value to the regression equation. Stepwise regression analysis is commonly used in behavioral research (Bakker, Van Der Zee, Lewig, & Dollard, 2006; Komarraju & Karau, 2005; Loveland, Gibson, Lounsbury, & Huffstetler, 2005; Wodarski, 1978).

The expected outcome of this process was a cluster of statistically significant predictors of potential for promotion.

Research Question Two *was explored next:* ***What factors within the HLI instrument are significantly related to promotion potential as rated by the subjects' supervisors?*** As in Research Question One's processes, a check was conducted for any redundant factors through correlation analysis. Stepwise multiple regression analysis was then conducted using the Potential for Promotion scores from the supervisor performance ratings as the dependent variable and the HLI assessment scores as possible explanatory variables to determine the extent to which they predict promotional potential. The desired outcome was a cluster of statistically significant predictors of leadership potential.

Legal Compliance

The guidelines published by The Society for Industrial and Organizational Psychology (*Principles for the validation and use of personnel selection procedures*, 2003) state that they are not intended to interpret federal, state, local, or case law regarding employee selection, but the guidelines can inform decision-making related to them. The focus of the entire publication is the validation of selection tools to ensure job-relatedness, as required by law. The guidelines suggest that certain "sources of evidence" help ensure validity, three of which are predictive relationship between selection method and on-the-job performance, job content relatedness, and internal structure of the test. Predictive validity is the first "source of evidence" offered by the guidelines; however it can be a difficult and time-consuming process for a practitioner to undergo, especially in the midst of other work priorities.

The Equal Employment Opportunity Commission's web site also provides guidelines regarding selection processes (http://www.uniformguidelines.com/uniformguidelines.html#59):

> The degree of relationship between selection procedure scores and criterion measures should be examined and computed, using professionally acceptable statistical procedures. Generally, a selection procedure is considered related to the criterion, for the purposes of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the 0.05 level of significance, which means that it is sufficiently high as to have a probability of no more than one (1) in twenty (20) to have occurred by chance. Absence of a statistically significant relationship between a selection procedure and job performance should not necessarily discourage other investigations of the validity of that selection procedure.

The methods by which these data were analyzed complied with both sets of guidelines.

Data Comparison

Once these steps were completed, a correlation analysis was conducted between the HLI assessment scores and the performance evaluations to answer ***Research Question Three: How do the performance evaluation predictors compare to the HLI instrument predictors?*** In this step, the researcher compared the performance evaluations' predictors of leadership potential identified in Research Question One with the assessment-based predictors identified in Research Question Two. To do this, the factors found to be most predictive statistically for each instrument were reviewed in relationship, content, and theory. The items found to be most for the instruments are compared and discussed in Chapters Four and Five.

Research Question Four: What other elements of the assessment instrument are significantly related to leadership performance ratings? The final research question examined other points of interest regarding the two data sets. For example, the Kruskal-Wallis analysis of

variance was used to determine if the subject's performance evaluation or HLI scores varied significantly depending upon level of management, years of service, or time in position (Salkind, 2004). Additional analysis was also conducted in this step including further analysis of the correlations between all possible variables, the Critical Thinking factor, examination of the halo effect, and principal component analysis to identify predictors of low promotion potential.

Summary

The answers to the research questions described within this methodology chapter will be presented in Chapter Four, as are additional *post hoc* analyses. Implications and recommendations developed based upon these results will be provided in Chapter Five.

CHAPTER IV - RESULTS

Introduction

The purpose of this chapter is to describe the results of the study based upon the methodological steps taken in Chapter Three. The chapter is organized in the following manner: First, a description of the process which reduced the original sample population from 375 to 175 will be provided. Secondly, characteristics of the sample population are summarized. The data analysis and results for each research question are then reviewed with supporting tables and figures. A summary of the findings concludes Chapter Four.

Data Analysis of Sample

Of the 375 managers who took the HLI assessment, 175 of their supervisors did not complete performance evaluations. Table 1 shows the breakdown by hospital.

Table 1.

HLI Assessment and Supervisor Ratings by Hospital

Hospital #	HLI Assessment	%	Supv Ratings	%	% HLI with Sup Ratings
1	20	5.33%	10	5.00%	50%
2	37	9.87%	-	-	0%
3	2	0.53%	-	-	0%
4	23	6.13%	20	10.00%	87%
5	56	14.93%	43	21.50%	77%
6	8	2.13%	5	2.50%	63%
7	29	7.73%	-	-	0%
8	17	4.53%	6	3.00%	35%
9	41	10.93%	37	18.50%	90%
10	34	9.07%	34	17.00%	100%
11	3	0.80%	3	1.50%	100%
12	14	3.73%	-	-	0%
13	46	12.27%	-	-	0%
14	17	4.53%	15	7.50%	88%
15	3	0.80%	3	1.50%	100%
16	25	6.67%	24	12.00%	96%
Total	**375**	**100%**	**200**	**100%**	**N/A**

As Table 1 shows, hospitals 2, 3, 7, 12, and 13 did not provide any supervisory ratings at all. Additionally, hospitals 1, 4, 5, 8, 9, 14, and 16 provided partial data. Three hospitals (10, 11 and 15) provided supervisory ratings for all test takers.

The hospitals which did not provide supervisory ratings were excluded from this study due to the lack of the Potential for Promotion score that was needed for use as the dependent variable in the multiple regression analysis. As an explanation for the missing data, the test creator from The Kingwood Group provided the following statement, “Some of the hospitals opted not to provide the performance evaluation information, and the response rate varied for supervisors who worked in hospitals which did approve of the performance evaluation process.” Additionally, five subjects had been employed or were in their current positions for fewer than 90 days; so their data were excluded from the study as well. Table 2 provides detailed sample information by hospital, including the removed and retained *n* sizes, and size and location of each hospital. Size was not provided for Hospital 9 in Michigan.

Table 2.

Removed and Retained Population including Hospital Size and Location

Hospital	Removed n	Retained n	Total	# Beds	State
1	10	10	20	91	OH
2	37	0	37	500	OH
3	2	0	2	303	OH
4	4	19	23	419	LA
5	14	42	56	220	OH
6	3	5	8	150	WV
7	29	0	29	124	IN
8	11	6	17	124	IL
9	5	36	41	--	MI
10	2	32	34	172	MI
11	0	3	3	91	OH
12	14	0	14	519	PA
13	46	0	46	389	KY
14	2	15	17	76	KS
15	0	3	3	320	KS
16	1	24	25	104	NC
Total	180	195	375	-	-

Of the 195 remaining records, 15 subjects were missing one supervisor rating score as shown in Table 3.

Table 3.

Missing Supervisor Ratings

Factor	Missing *n*
Supervisor Ranking	8
Overall Performance	4
Conscientiousness	2
Drive for Results	1
Total	**15**

An acceptable standard practice is to replace missing data points with the mean score of all other subjects for that variable, when the missing data constitute less than fifteen percent of the variable data (George & Mallery, 2006). This procedure was employed to populate the 15 omitted fields.

In addition, the data set did not contain Critical Thinking scores for 56 of the remaining 195 subjects. The Kingwood Group stated that while the Critical Thinking instrument is part of the HLI assessment, a separate step is required to take that portion. They attribute the missing data to the response rate of the subjects in taking that separate step. As opposed to the missing supervisor performance data, the data for subjects with missing Critical Thinking scores were retained. In the data analysis in which Critical Thinking scores were used, only the subjects with Critical Thinking scores were included, and the results are annotated with "*n=139."

A summary of the reduced sample by hospital is provided next in Table 4.

Table 4.

Data by Hospital

Hospital #	HLI Assessment	Critical Thinking	Supv Ratings
1	10	9	10
4	19	14	19
5	42	32	42
6	5	4	5
8	6	-	6
9	36	30	36
10	32	28	32
11	3	-	3
14	15	-	15
15	3	2	3
16	24	20	24
Total	**195**	**139**	**195**

Table 4 shows that four hospitals (5, 9, 10, and 16) provided 68.7 percent of the sample size for the HLI and Supervisory Ratings. The other hospitals provided this information on between three and 19 subjects. A one-way analysis of variance (ANOVA) was conducted between the hospitals and the dependent variable (Potential for Promotion) to determine the differences between hospitals in the dependent variable score. The differences between hospitals were not significant (p=.984). Therefore, all of the hospitals' data were used in this study.

It is important to examine the removed data sets in comparison to the retained data set to ensure that the changes made in preparing the data for analysis did not substantively alter the population's original distribution. Therefore, an Independent Samples t-test was conducted on the reduced sample (n=195) as compared to those who were removed from the sample (n=180) for Age, Gender, Race, Years of Service, Time in Position, and Leadership Level. This test is used to compare the sample means when "two samples share some variable of interest in common, but there is no overlap in membership of the two groups" (George & Mallery, 2006, p. 134). The test statistics for all groupings are shown below in Table 5.

Table 5.

Independent Samples T-test for Removed and Retained Hospitals

		Levene		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Diff	Std. Error Diff	95% Confidence Interval of the Difference	
									Lower	Upper
Gender	Equal var	1.56	.212	.62	347	.533	.028	.046	-.061	.118
	Unequal var			.62	346.63	.533	.028	.045	-.061	.118
Age	Equal var	9.16	.003	1.50	347	.135	.068	.046	-.021	.158
	Unequal var			1.50	346.78	.133	.068	.045	-.021	.158
Race	Equal var	12.03	.001	1.70	347	.089	-.037	.022	-.080	.006
	Unequal var			1.73	308.60	.085	-.037	.021	-.079	.005
Emp Duration	Equal var	.09	.766	.35	347	.728	.033	.096	-.156	.222
	Unequal var			.35	341.40	.729	.033	.096	-.156	.223
Pos Duration	Equal var	3.02	.083	-.56	347	.576	-.074	.131	-.332	.185
	Unequal var			-.56	329.21	.578	-.074	.132	-.334	.186
Ldrship Level	Equal var	.012	.912	1.96	347	.051	.217	.111	-.001	.435
	Unequal var			1.96	346.49	.051	.217	.111	-.001	.435

As shown in Table 5, based upon the Levene Test, Equality of Variances was either assumed or not assumed. These p scores all indicate that there is not a significant difference between the two

subpopulation samples, although the p score for leadership level was just below the significance cutoff at .051. It is important to note that, because the ethnic makeup of the sample was predominantly Caucasian, the race variables were reduced to two categories (Caucasian and non-Caucasian) for analysis. Tables 6 and 7 show descriptive data for the two groupings (data removed and retained) prior to the recoding.

Table 6.

Descriptive Statistics for Race – Removed Data Set

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	White/Caucasian	175	97.2	97.8	97.8
	Black/African American	2	1.1	1.1	98.9
	Hispanic/Latino American	1	.6	.6	99.4
	American Indian/Alaskan Native	1	.6	.6	100.0
	Total	179	99.4	100.0	
Missing	System	1	.6		
Total		180	100.0		

Table 7.

Descriptive Statistics for Race – Retained Data Set

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	White/Caucasian	182	93.3	93.3	93.3
	Black/African American	8	4.1	4.1	97.4
	Hispanic/Latino American	1	.5	.5	97.9
	Asian/Pacific Islander	1	.5	.5	98.5
	Other	1	.5	.5	99.0
	rather not say	2	1.0	1.0	100.0
	Total	195	100.0	100.0	

Tables 6 and 7 show that the retained data set contained 4.1 percent Black/African American subjects (*n*=8) as compared to 1.1 percent (*n*=2) in the removed data set. Additionally, the removed data set contained one American Indian/Alaskan Native subject; and the retained data set had none. One Hispanic/Latino American was also removed from the sample. The White/Caucasian population for the removed sample is 97.2 percent as opposed to 93.3 percent for the retained sample. The differences in the two data sets are slightly statistically significant, but it would have been more disturbing had the removed data set contained a larger percentage of protected classifications. Instead, the remaining data set actually has a smaller percent of White/Caucasian subjects (usually the dominant population in leadership studies in the United States). In other words, the removal process does not appear to have been biased against protected classifications of employees.

Characteristics of the Sample

In Table 8, selected characteristics of the sample population are described.

Table 8.

Selected Characteristics of the Sample Population

Characteristic	Number	Percent
Sample Total	195	100%
Gender		
Male	48	24.6
Female	145	74.4
Rather Not Say	2	1.0
Age		
Under 40	56	28.7
40 or Over	138	70.8
Rather Not Say	1	.5
Race or Ethnic Background		
White/Caucasian	182	93.3
Black/African American	8	4.1
Hispanic/Latino American	1	.5
Asian/Pacific Islander	1	.5
Other	1	.5
Rather Not Say	2	1.0

Employment Duration		
3-6 months	4	2.1
6-12 months	1	.5
1-2 years	21	10.8
2-5 years	37	19.0
5+ years	132	67.7
Position Duration		
3-6 months	9	4.6
6-12 months	11	5.6
1-2 years	42	21.5
2-5 years	59	30.3
5+ years	74	37.9
Leadership Level		
Charge Nurse	5	2.6
Supervisor	36	18.5
Middle Manager	71	36.4
Director	55	28.2
Vice President	14	7.2
Other	14	7.2

Table 8 shows that there were more than three times as many women than men in this study and almost 2.5 times as many subjects 40 years of age or older than the under-40 population. Racially, they were mostly Caucasian. Most of the population had employment durations of more than two years, with 67.7 percent having five or more years of service. Similarly, most of the population had been in the current positions for at least one year, with 37.9 percent in their positions for five or more years. The subjects were fairly evenly distributed among the management levels, with almost 65 percent of them belonging to the middle manager or director ranks. “Rather not say” responses were identified as discrete missing variables. The “Other” response for race was coded as non-Caucasian because the researcher assumed that one who

would select “Other” would be non-Caucasian. “Other” responses for Leadership Level were coded as discrete missing variables.

The Independent Variables

Table 9 presents descriptive statistics for the supervisor performance ratings, indicating a low mean of 3.08 for Ranking and a high of 5.61 for Conscientiousness. Ranking is based upon a different scale from the other factors, so it does not lend itself to mean comparison as readily as the other factors, although it has been included in the table for informational purposes. Excluding ranking, the lowest mean score is 4.83 for promotion potential. Skewness for the factors ranged from -.541 to .028; and Kurtosis ranged from -.749 to .636. Because these scores were well within the threshold of ±1, all of the variables have “excellent” symmetric and peak characteristics (George & Mallery, 2006, p. 99).

Table 9.

Supervisor Performance Evaluation Scores

Factor	Mean	Median	SD	Skew	Kurtosis
Drive for Results	5.18	5.00	1.057	-.365	.636
Conscientiousness	5.61	6.00	1.032	-.541	.284
Customer Orientation	5.35	5.00	1.061	-.199	-.399
Emotional Evenness	4.99	5.00	1.260	-.371	-.309
Innovative	4.81	5.00	1.118	-.100	-.254
Multi-Tasking	5.08	5.00	1.081	-.165	-.181
Openness to Change	5.09	5.00	1.073	-.428	.358
Self Development	5.10	5.00	1.065	-.284	-.139
Self Confidence	5.16	5.00	1.112	-.342	-.749
Promotion Potential	4.83	5.00	1.183	-.253	.136
Overall Performance	5.17	5.00	.951	-.491	.162
Ranking	3.08	3.00	1.210	.028	-.583

Table 10 presents the descriptive statistics for the HLI Assessment scores.

Table 10.

HLI Assessment Scores

Factor	Mean	Median	SD	Skew	Kurtosis
Achievement	56.7846	57.0000	5.48614	-.045	-.226
Conscientiousness	59.5282	60.0000	4.91715	-.537	.819
Customer Orientation	52.3282	52.0000	5.05587	-.466	.881
Emotional Evenness	43.3641	44.0000	5.35307	-.544	.086
Innovative	38.1487	38.0000	5.00551	-.151	-.303
Multi-Tasking	44.1692	45.0000	4.97800	.102	.375
Openness to Change	32.1231	32.0000	3.51803	-.517	1.784
Self Development	42.0205	43.0000	4.14301	-.532	.827
Self Confidence	51.8154	52.0000	5.17695	.000	-.287
Critical Thinking*	39.83	41.00	6.588	-.678	.610

*n=139

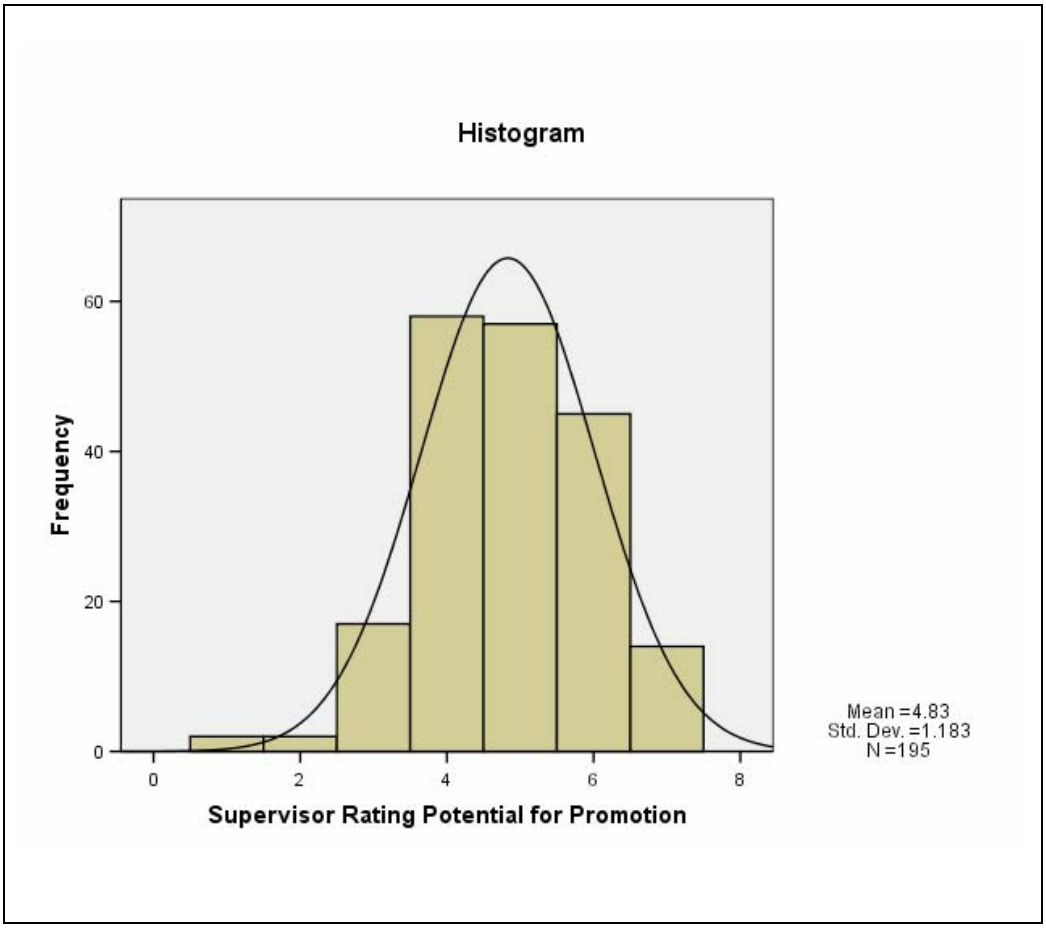
Table 10 shows that the HLI mean scores ranged from 39.83 for Critical Thinking to 59.53 for Conscientiousness. The Critical Thinking factor is based upon a different scale from the other factors, so it does not easily lend itself to comparison with the other factors. The lowest mean score after Critical Thinking was Self Development at 42.02, with a range between the low and high means of 17.51. Skewness ranged from -.678 to .102; and Kurtosis ranged from -.287 to 1.784. All scores fell within the “excellent” threshold (±1) for normal distribution except Openness to Change, which fell into the “acceptable” threshold (±2) (George & Mallery, 2006, p. 99), especially given that the skewness for the Openness to Change factor is well under ± 1.0.

The Dependent Variable

Figure 4 provides a visual representation of the normal distribution curve of Promotion Potential Supervisor Rating. This rating was spotlighted due to its importance in this study as the dependent variable.

Figure 4.

Histogram of Supervisor Rating - Potential for Promotion



Although the distribution as shown in Figure 4 appeared to be normal, Kolmogorov-Smirnov (K-S) and Shapiro-Wilk tests for normality were conducted to ensure normal distribution. The results are shown in Table 11.

Table 11.

K-S and Shapiro-Wilk Tests for Normality

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Supervisor Rating Potential for Promotion	.164	195	.000	.927	195	.000

The statistical results displayed in Table 11 show a low significance score for both the K-S and Shapiro-Wilk tests, which indicates that the Potential for Promotion scores are not normally distributed, although they appear to be in the histogram in Figure 8. The result can often be caused by skew and kurtosis, but that is not true in this case; Table 9 shows that both are acceptable for the Potential for Promotion variable. A “stem and leaf” display helps to bring the problem into view, as shown in Figure 5.

Figure 5.

Stem and Leaf Display – Potential for Promotion Variable

```
Supervisor Rating Potential for Promotion Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2.00        1 .  0
      .00        1 .
     2.00        2 .  0
      .00        2 .
    17.00        3 .  00000000
      .00        3 .
    58.00        4 .  00000000000000000000000000000000
      .00        4 .
    57.00        5 .  00000000000000000000000000000
      .00        5 .
    45.00        6 .  000000000000000000000
      .00        6 .
    14.00        7 .  0000000

 Stem width:         1
 Each leaf:       2 case(s)
```

Figure 5 shows a very light distribution on the lower end of the scale which then quickly ramps up from the “3” score with 17 entries to the “4” score with 58 scores. However, the mean and standard deviation for this factor’s scores are perfectly acceptable. GraphPad, a software company specializing in data analysis, provides a web site as a resource for its users. On this web site, the efficacy of normality tests is discussed within the context of the Kolmogorov-Smirnov test ("How useful are normality tests?" 2006). The site quotes R.B. D’Agostino as saying, “The Kolmogorov-Smirnov test is only a historical curiosity. It should never be used” (http://www.graphpad.com/library/BiostatsSpecial/article_197.htm). D’Agostino believes that the test is too simple and does not discriminate the data well. A second test for normal distribution, the Shapiro-Wilk test, can also be problematic, especially when variable scores are repeated, which is the case with the current sample. Because the only potential objection to the normality of the data set was an analysis technique that is in question, it was treated as if it were

normally distributed, and underwent all applicable parametric tests that would be conducted on a normally distributed data set.

For this study, the researcher opted to combine the "Poor" and "Well Below Average" scores into one category because each had only one entry. Although this adjustment did not significantly improve the Kolmogorov-Smirnov score, the data set did appear to be more normally distributed afterwards.

It is also important to note that some consideration was given to mathematically combining the Potential for Promotion scores with other supervisor ratings such as Overall Performance or Supervisor Ranking. This was contemplated due to the limited whole number range (1-7) of the Potential for Promotion scores. However, it was ultimately decided that Potential for Promotion was the key dependent variable; the addition of other variables would dilute the importance of the construct of promotional potential.

Bias Check of the Dependent Variable

A One-Way Analysis of Variance showed no statistically significant effects for gender, race, or age. In other words, the researcher found no gender, race, or age biases within the dependent variable. Additionally, no main or interaction effects were found for these demographic characteristics.

Once the normal distribution of the population and variables was established, the data sets were "advanced" to the next step in the process: addressing the research questions.

Research Question One – Data Analysis and Results

What factors within the performance evaluation instrument are significantly related to promotion potential as rated by the subjects' supervisors?

As described in Chapter Three, a correlation analysis was first conducted to identify any factors which correlated too closely with other variables, as shown in Table 12.

Table 12. Correlation Analysis - Performance Evaluation (n = 195)

		Supervisor Rating Drive for Results	Supvsr Rating Conscientious	Supvsor Rating Custmr Orientn	Supvisor Rating Emot Evens	Supervsr Rating Innovative	Supervsr Rating Multi-Tasking	Supervsr Rating Open to Change	Supervsr Rating Self Developt	Supervsr Rating Self Confid	Supvsor Rating Promo Potential	Supervsr Rating Overall Perform	Supervsr Ranking
Supervisor Rating Drive for Results	Pearson Correlation	1											
Supervisor Rating Conscientiousness	Pearson Correlation	.552(**)	1										
Supervisor Rating Customer Orientation	Pearson Correlation	.481(**)	.491(**)	1									
Supervisor Rating Emotional Evenness	Pearson Correlation	.396(**)	.354(**)	.515(**)	1								
Supervisor Rating Innovative	Pearson Correlation	.592(**)	.371(**)	.419(**)	.452(**)	1							
Supervisor Rating Multi-Tasking	Pearson Correlation	.583(**)	.501(**)	.410(**)	.496(**)	.585(**)	1						
Supervisor Rating Openness to Change	Pearson Correlation	.559(**)	.390(**)	.489(**)	.538(**)	.723(**)	.589(**)	1					
Supervisor Rating Self Development	Pearson Correlation	.533(**)	.389(**)	.369(**)	.350(**)	.480(**)	.539(**)	.533(**)	1				
Supervisor Rating Self Confidence	Pearson Correlation	.497(**)	.329(**)	.393(**)	.560(**)	.526(**)	.555(**)	.524(**)	.443(**)	1			
Supervisor Rating Potential for Promotion	Pearson Correlation	.701(**)	.498(**)	.524(**)	.545(**)	.626(**)	.696(**)	.633(**)	.562(**)	.640(**)	1		
Supervisor Rating Overall Performance	Pearson Correlation	.718(**)	.631(**)	.620(**)	.612(**)	.623(**)	.699(**)	.652(**)	.584(**)	.681(**)	.773(**)	1	
Supervisor Ranking	Pearson Correlation	-.604(**)	-.529(**)	-.523(**)	-.473(**)	-.450(**)	-.569(**)	-.501(**)	-.518(**)	-.549(**)	-.661(**)	-.706(**)	1

** Correlation is significant at the 0.01 level (2-tailed).

Table 12 shows the correlations among the supervisor rating factors. Absolute values of the correlations ranged from r=.329 (Self Confidence and Conscientiousness) to r=.773 (Overall Performance and Potential for Promotion). All correlations were significant at the $p<.01$ level. The intent of this step was to eliminate any variables that were too strongly correlated with other variables, and which were effectively duplicative. Salkind (2004) provides a rule of thumb for interpreting correlation coefficients assuming a significance level ($p<.05$) that is summarized in Table 13 (p. 88):

Table 13.

Interpreting a Correlation Coefficient

Size of the Correlation Coefficient	**General Interpretation**
.8 to 1.0	Very Strong Relationship
.6 to .8	Strong Relationship
.4 to .6	Moderate Relationship
.2 to .4	Weak Relationship
.0 to .2	Weak or No Relationship

Based on Table 13, factors that correlate with other factors at r=.8 or higher should be eliminated from further analysis. In fact, George and Mallery (2006) stated that “variables that correlate higher than r=.5 should be scrutinized carefully before both are included in a regression analysis” (p. 195) because the multiple regression results could be compromised otherwise. While the absolute correlation coefficient values ranged from r=.329 (weak relationship) to r=.773 (strong relationship), none exceeded the r=.8 threshold. High correlations were expected

in this step because all of the variables derive from the same source. So, the r=.5 threshold suggested by George and Mallery was not upheld for any variable with even one correlation that exceeded r=.5. However, every possible correlation for the Overall Performance factor yielded an r of greater than r=.5. This called for it to be removed from further analysis. Additionally, the researcher elected to remove the Supervisor Ranking variable from further analysis because it does not describe a specific characteristic, and it varied closely with Promotion Potential.

Next, stepwise multiple regression analysis was conducted using the Potential for Promotion performance evaluation score as the dependent variable, and the ten remaining performance factors as independent variables. Multiple regression analysis is one of the most common approaches to show the influence of two or more variables on the dependent variable. The stepwise method combines both forward and backward methods, which removes variables that weaken other more qualified variables. Stepwise is the most frequently used of the regression models (George & Mallery, 2006). Table 14 presents the results of the stepwise multiple regression analysis.

Table 14.

Stepwise Multiple Regression Analysis – Performance Ratings (n = 195)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F Change	Sig F Change
1	.701(a)	.491	.489	.825	186.439	.000
2	.785(b)	.616	.612	.718	62.628	.000
3	.817(c)	.668	.662	.670	29.395	.000
4	.826(d)	.682	.675	.657	8.714	.004
5	.830(e)	.689	.681	.652	4.134	.043

a Predictors: (Constant), Supervisor Rating Multi-Tasking

b Predictors: (Constant), Supervisor Rating Multi-Tasking, Supervisor Rating Drive for Results

c Predictors: (Constant), Supervisor Rating Multi-Tasking, Supervisor Rating Drive for Results, Supervisor Rating Self Confidence

d Predictors: (Constant), Supervisor Rating Multi-Tasking, Supervisor Rating Drive for Results, Supervisor Rating Self Confidence, Supervisor Rating Openness to Change

e Predictors: (Constant), Supervisor Rating Multi-Tasking, Supervisor Rating Drive for Results, Supervisor Rating Self Confidence, Supervisor Rating Openness to Change, Supervisor Rating Customer Orientation

The Model Summary in Table 14 and the notes below it present the five models derived from the stepwise regression analysis which included Multi-Tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation. Excluded variables were

Conscientiousness, Emotional Evenness, Innovative, and Self Development, because they did not sufficiently add to the regression model. The first model which included only Multi-Tasking had an R of .701, which indicates a substantial correlation between the factor and the independent variable, Potential for Promotion. In fact, more than 49 percent of the variability of the dependent variable can be explained by the Drive for Results rating. These types of strong correlations between the performance factors were expected because the performance ratings and the potential for promotion were all provided by the same source. The Model Summary shows that a factor was added in each successive model, incrementally increasing the R and R-Square scores, culminating in five factors achieving an R of .830 and Adjusted R-Square of .681, meaning that almost 70 percent of the variance of the Potential for Promotion scores can be explained by the variation of those five factors. All results were significant to the $p<.05$ level. The Durbin-Watson test for independent errors yielded a score of 1.903. Because "The test statistic can vary between 0 and 4 with a value of 2 meaning the residuals are uncorrelated" (Field, 2005, p. 730), the independent errors were found to be acceptable.

The results of the stepwise multiple regression process is a number of models which combine the independent variables in such a way as to best explain the variability in the dependent variable. Stepwise regression is complete when variables added to the equation no longer make a significant contribution to the model. In this particular analysis, the SPSS program calculated five independent variables which contributed to the model. Therefore, the five factors included in the fifth model (Multi-Tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation) were selected for comparison with the HLI predictors in Research Question Three. The Variance Inflation Factor (VIF) test for multicollinearity for this model were 1.943 (Multi-Tasking), 1.845 (Drive for Results), 1.647 (Self Confidence), 1.912

(Openness to Change), and 1.452 (Customer Orientation). Field (2005) cites Myers (1990) as stating that a value of 10 or lower is acceptable. However, he also cites Bowerman and O'Connell (1990) as stating that multicollinearity might exist if the average VIF is greater than one. Because the numbers were well under 10 and just over one, the researcher assumed that multicollinearity was not biasing the model. Standardized residuals statistics were acceptable at -2.664 (minimum) and 2.535 (maximum), with a mean of 0.00 (Field, 2005).

Research Question Two – Data Analysis and Results

What factors within the HLI instrument are significantly related to promotion potential as rated by the subjects' supervisors?

The first step in answering Research Question Two was to conduct bi-variate correlation analysis on the HLI assessment scores. Just as in Research Question One, the purpose was to eliminate any variables that were too strongly correlated with other variables, and which were effectively redundant. Table 15 shows the correlation results of the HLI assessment scores.

Table 15. Correlation Analysis - HLI Scores

		HLI Achvmnt	HLI Cust Orientation	HLI Conscien	HLI EmotEven	HLI Innovative	HLI Multi-Tasking	HLI Open to Change	HLI Self Confid	HLI Self Devlpmnt	Critical Thinking
HLI Achievement	Pearson Correlation	1									
	Sig. (2-tailed)										
	N	195									
HLI Customer Orientation	Pearson Correlation	.500(**)	1								
	Sig. (2-tailed)	.000									
	N	195	195								
HLI Conscientiousness	Pearson Correlation	.593(**)	.515(**)	1							
	Sig. (2-tailed)	.000	.000								
	N	195	195	195							
HLI Emotional Evenness	Pearson Correlation	.314(**)	.486(**)	.325(**)	1						
	Sig. (2-tailed)	.000	.000	.000							
	N	195	195	195	195						
HLI Innovative	Pearson Correlation	.382(**)	.372(**)	.354(**)	.292(**)	1					
	Sig. (2-tailed)	.000	.000	.000	.000						
	N	195	195	195	195	195					
HLI Multi-Tasking	Pearson Correlation	.288(**)	.189(**)	.364(**)	.255(**)	.279(**)	1				
	Sig. (2-tailed)	.000	.008	.000	.000	.000					
	N	195	195	195	195	195	195				
HLI Openness to Change	Pearson Correlation	.426(**)	.385(**)	.533(**)	.439(**)	.535(**)	.433(**)	1			
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000				
	N	195	195	195	195	195	195	195			
HLI Self Confidence	Pearson Correlation	.550(**)	.584(**)	.540(**)	.502(**)	.573(**)	.379(**)	.531(**)	1		
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000			
	N	195	195	195	195	195	195	195	195		
HLI Self Development	Pearson Correlation	.476(**)	.434(**)	.374(**)	.322(**)	.472(**)	.261(**)	.469(**)	.438(**)	1	
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		
	N	195	195	195	195	195	195	195	195	195	
Critical Thinking Inventory Scale*	Pearson Correlation	-.094	-.134	.068	-.115	-.060	-.017	-.027	-.046	-.078	1
	Sig. (2-tailed)	.270	.116	.426	.177	.486	.843	.754	.593	.359	
	N	139	139	139	139	139	139	139	139	139	139

** Correlation is significant at the 0.01 level (2-tailed).

Table 15 shows the results of the bi-variate correlation analysis. The correlation absolute values ranged from .017 (Critical Thinking and Multi-Tasking) to .593 (Conscientiousness and Achievement). Based on the rule of thumb in Table 13, any factors correlating with other factors at .8 or higher would be eliminated from further analysis. The Critical Thinking correlation coefficients were not significant at $p<.05$, and so this factor was "safe" from elimination without even evaluating its correlation coefficients. The other factors' absolute correlation coefficient values (all significant at $p<.01$) ranged from .189 (weak to no relationship) to .593 (moderate relationship), but none of the coefficient scores for this data set exceeded the .8 threshold. Therefore, none of the HLI Assessment factors was eliminated from further analysis.

Next, stepwise multiple regression analysis was conducted using the Potential for Promotion performance evaluation score as the dependent variable, and the HLI assessment factors as independent variables as shown in Table 16.

Table 16.

Stepwise Multiple Regression Analysis – HLI Scores (n = 195)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.280(a)	.078	.071	1.142
2	.331(b)	.109	.096	1.126

a Predictors: (Constant), HLI Achievement

b Predictors: (Constant), HLI Achievement, HLI Openness to Change

c Dependent Variable: Supervisor Rating Potential for Promotion

The expected outcome was a cluster of statistically significant HLI Assessment predictors of potential for promotion. However, Table 16 shows that only two of the HLI assessment factors significantly predict variation of the independent variable (Potential for Promotion) at $p<05$ with this sample. In the first model, Achievement had an R of .280, indicating a weak relationship between it and the dependent variable. The Adjusted R-Square of .071 indicates that 7.1 percent of the variation in the dependent variable can be explained by the Achievement score. The second model couples the Openness to Change factor with Achievement, resulting in an R of .331 and an Adjusted R-Square of .096. It is important to note that individual personality variables in assessments do not tend to show the high correlation levels revealed when comparing the supervisory ratings to the Potential for Promotion rating, because they derive from the same source and would be expected to strongly correlate. For

instance, Barrick *et al.* (2003) conducted a meta-analysis on the relationship between the FFM and Holland's occupational types, and cited personality trait correlations of r=.20 and r=-.25 as significant and meaningful. As a matter of fact, Conscientiousness has been found to be "the only FFM construct to predict supervisory ratings of job performance across all jobs and organizations" (Witt, Burke, Barrick, & Mount, 2002, p. 164), yet the correlations cited to substantiate that claim "average in the low teens (e.g., r= .10 in Salgado, 1997)" (p. 165). In light of these comparisons to other research standards in the field, the Achievement and Openness to Change factors do show significant ability to predict the variation in the dependent variable; they will represent the HLI assessment as the best predictors of Promotion Potential.

The VIF and Tolerance scores for this model were acceptable at 1.209 and .827 respectively, showing low likelihood of multicollinearity. The Durbin-Watson test for independent errors yielded a score of 1.264, which was acceptable, although a score of closer to two would have been ideal. Standardized residuals statistics were acceptable at -2.616 (minimum) and 2.290 (maximum), with a mean of .064 (Field, 2005).

It is important to note that the regression analysis described above was programmed to exclude cases listwise. When the regression analysis was conducted using all of the HLI Assessment variables, the 56 cases that were missing Critical Thinking scores were entirely omitted from the analysis. When the stepwise regression analysis was conducted again using the "replace with mean" option, only Achievement was identified as a predictor ($R=.241$, $R^2=.053$), and Openness to Change was no longer included in the model. Furthermore, when the regression analysis was conducted again without including the Critical Thinking variable at all, Achievement once again was the only identified predictor. This outcome called for additional analysis: What is the difference between those subjects who did and did not take the Critical

Thinking portion of the HLI assessment? To answer this query, an independent samples t-test was conducted between the two groups for the demographic variables (age, gender, race, employment duration, position duration, and management level), as shown in Table 17.

Table 17.

Independent T-test for Critical Thinking and Demographic Variables

		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig. (2-tailed)
Gender	Equal variances assumed	1.342	.248	-.607	191	.545
	Not assumed			-.587	71.565	.559
Age	Assumed	1.933	.166	-.752	192	.453
	Not assumed			-.727	69.150	.470
Race	Assumed	1.541	.216	.629	193	.530
	Not assumed			.573	65.743	.569
Employment Duration	Assumed	11.113	.001	1.991	193	.048
	Not assumed			2.511	119.871	.013
Position Duration	Assumed	2.140	.145	1.990	193	.048
	Not assumed			2.219	90.881	.029
Mgmt Level	Assumed	1.312	.254	1.153	179	.250
	Not assumed			1.107	65.927	.272

As Table 17 shows, the t-test yielded differences among the two groups (did and did not take the Critical Thinking portion of the HLI Assessment) for Employment and Position Duration. Table 18 shows the means of the two groupings.

Table 18.

Mean Comparison of Employment and Position Duration for Critical Thinking Score

	Took Critical Thinking	N	Mean	Std. Deviation	Std. Error Mean
Employment Duration	No	46	5.72	.584	.086
	Yes	149	5.43	.925	.076
Position Duration	No	46	5.20	.934	.138
	Yes	149	4.83	1.149	.094

As Table 18 shows, those with longer Employment or Position Duration were less likely to have taken the Critical Thinking Inventory.

The researcher then conducted a similar t-test to compare the means of the HLI Assessment variables of those who did and did not take the Critical Thinking portion of the HLI. Significant differences existed for Customer Orientation, Emotional Evenness, Openness to Change (an identified predictor), Self Confidence (an identified predictor), and Self Development. There were not significant differences among the groups for Achievement, Conscientiousness, Innovative, and Drive for Results (an identified predictor). Interestingly, the mean scores were higher for those who did take the Critical Thinking portion of the HLI for

every single HLI variable. The full meaning of these differences is unclear; however, it is apparent that there is some systematic variation that is affecting the outcome of the regression analysis.

Research Question Three – Data Analysis and Results

How do the performance evaluation predictors compare to the HLI instrument predictors?

In this section, the predictors of leadership potential from supervisor ratings identified in Research Question One are compared with the HLI assessment-based predictors identified in Research Question Two. The factors found to be most predictive statistically for each instrument are reviewed in more detail pertaining to their relationships, background, and related theories. As discussed in the previous Research Question Sections, the HLI predictors identified were Achievement and Openness to Change, while the supervisor rating predictors were Multi-Tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation (shown in Table 19). The implications of the differences and overlap of the predictors will be discussed at length in Chapter Five. This section will address how the two sets of predictors compare and contrast.

Table 19.

HLI and Supervisory Rating Predictors

HLI Predictors	Supervisor Rating Predictors
Achievement	Multi-Tasking
Openness to Change	Drive for Results
	Self Confidence
	Openness to Change
	Customer Orientation

About the Variables

The comparison begins with additional detail on each of the variables.

The HLI Assessment Variables

The HLI predictors of Achievement and Openness to Change both have a basis in the Big Five Personality Theory as overviewed in Chapter Two of this dissertation. The Achievement scale on the HLI is based upon a subset of the Conscientious factor of the Big Five Personality Theory (J. E. Smith, 2006). The Openness to Change scale on the HLI is based upon a subset of the Big Five's Openness factor. The Kingwood Group describes these two factors in its Selection Report, summarized in Table 20.

Table 20.

Kingwood Descriptions of Achievement and Openness to Change - HLI

Factor	**General Description**	**Low Score Description**	**High Score Description**
Achievement Orientation	Desire to accomplish career aspirations and goals.	Lacking ambition and drive for success. Aimless.	Desires to be successful. Wants to reach long term career goals.
Openness to Change	Adapts to changing priorities, demands, and work processes.	Resists change. Comfortable with *status quo*.	Adjusts to changing demands & priorities and encourages others to do the same.

Table 20 shows that the Achievement Orientation includes characteristics such as accomplishment, goal-setting, ambition, and success, while Openness to Change includes characteristics such as Adaptability and Encourages Others.

The Supervisor Rating Form Variables

The Supervisor Rating form, including the categories of Multi-tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation, was developed by the Kingwood Group for purposes of assessment validation. Consequently, the bulk of the rating categories mirror the HLI Assessment categories. For instance, the Supervisor Rating categories

of Multi-Tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation were derived from the HLI Achievement Orientation scale. The descriptions of the top predictive factors (as determined in Research Question One) are provided in the Performance Evaluation form, and are summarized below in Table 21.

Table 21.

Kingwood Descriptions of Predictive Supervisory Ratings

Factor	**General Description**	**Low Score Description**	**High Score Description**
Multi-Tasking	Ability to work on multiple tasks and meet multiple demands.	Easily frustrated by multiple demands. Can only focus on one thing at a time.	Prefers working on multiple tasks. Easily juggles multiple tasks. Thrives on multiple demands.
Drive for Results	Desire to be successful and reach both business and personal career goals.	Content with career status. Apathetic about success. Fails to set business goals.	Ambitious. Goal oriented. Success-oriented.
Self Confidence	Belief in own capabilities and having a positive self-image.	Low self-esteem. Has trouble making decisions. Unsure.	Competent. Optimistic. Enthusiastic.

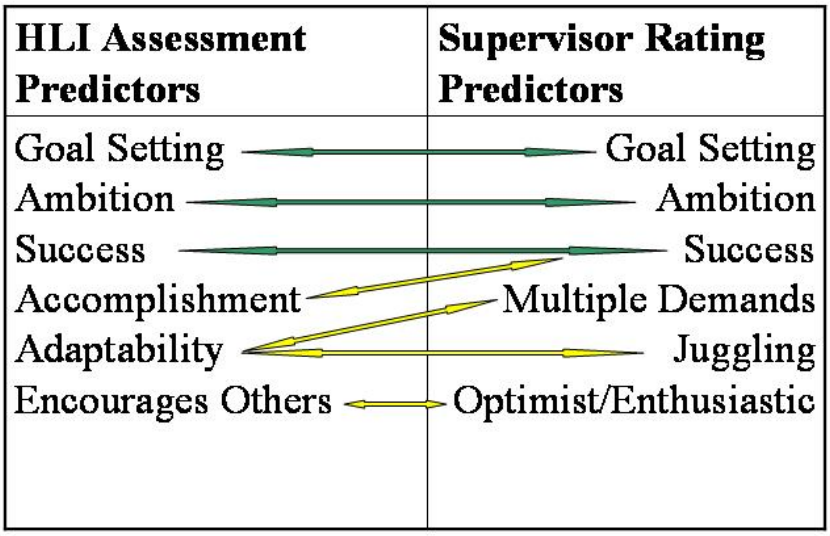
Openness to Change	Willingness to try new or different methods or approaches.	Inflexible. Resists Change. Complains when asked to adjust priorities.	Quickly adjusts to change. Flexible to changes in priorities and demands.
Customer Orientation	Attentiveness to the needs of both internal and external customers.	Uncaring. Sees customers as an imposition. Often sullen and unfriendly.	Courteous. Responsive. Service-oriented. Customer-driven.

Table 21 shows that Multi-Tasking references multiple demands and tasks and juggling them. Drive for Results factor includes such characteristics as goal-setting, success, and ambition while Self Confidence refers to a positive self image, optimistic and enthusiastic. Openness to Change focuses on adaptability, and Customer Orientation references a focus on providing service to internal and external customers.

Using the key descriptors from each of the identified predictors, a comparison of the predictors from both instruments is shown in Figure 6.

Figure 6.

Key Characteristics Comparison



The comparisons set forth in Figure 6 show a great deal of symmetry of the constructs between the predictors of the two instruments. Goal Setting, Ambition, and Success are the explicitly shared characteristics in the descriptions of the predictors. Accomplishment could also be compared to Success favorably. The characteristic of Adaptability could also compare favorably to Multiple Demands and the term "Juggling" that was used in the supervisor rating form, presumably intended to mean, "To keep more than two activities in progress at one time" (*The American Heritage Dictionary*, 2001, p. 462). One who is optimistic and enthusiastic tends to encourage others, so this is also a connection which can be made.

Quantitative Analysis of the Variables

Using the five identified predictors, a correlation analysis was conducted to evaluate the factors' relationships with each other, as shown in Table 22.

Table 22.

HLI and Supervisor Rating Predictor Correlations

		HLI Achievmnt	HLI Open to Change	Sup Rat Drive for Results	Sup Rat Customer Orientation	Sup Rat Multi-Tasking	Sup Rat Open to Change	Sup Rat Self Confidence
HLI Achievement	Pearson Correlation	1						
	Sig. (2-tailed)							
	N	195						
HLI Open to Change	Pearson Correlation	.426**	1					
	Sig. (2-tailed)	.000						
	N	195	195					
Sup Rat Drive for Results	Pearson Correlation	.366**	.195**	1				
	Sig. (2-tailed)	.000	.006					
	N	195	195	195				
Sup Rat Customer Orientation	Pearson Correlation	.229**	.093	.481**	1			
	Sig. (2-tailed)	.001	.195	.000				
	N	195	195	195	195			
Sup Rat Multi Tasking	Pearson Correlation	.204**	.164**	.583**	.410**	1		
	Sig. (2-tailed)	.004	.022	.000	.000			
	N	195	195	195	195	195		
Sup Rat Open to Change	Pearson Correlation	.253**	.184	.559**	.489**	.589**	1	
	Sig. (2-tailed)	.000	.010	.000	.000	.000		
	N	195	195	195	195	195	195	
Sup Rat Self Confidence	Pearson Correlation	.170**	.161*	.497**	.393**	.555**	.524**	1
	Sig. (2-tailed)	.017	.025	.000	.000	.000	.000	
	N	195	195	195	195	195	195	195

** Correlation is significant at the 0.01 level (2-tailed).

Table 22 presents the results of the correlation analysis between the HLI and Supervisor Rating predictors of promotional potential. All of the correlations are significant at the *p*=.05 level, and range from r=.161 (HLI Openness to Change and Supervisor Rating Self Confidence) to r=.583 (Supervisor Ratings Multi-Tasking and Drive for Results). As expected based upon their strong correlations from the multiple regression analysis in Research Question One, the Pearson Correlation for the supervisory ratings carried the highest correlations ranging from r=.481 to r=.583 with each other and r=.170 to r=.366 with the HLI factors of Achievement and Openness to Change. None of the factors was so closely correlated to the other as to be duplicative, although Supervisor Ratings of Drive for Results/Multi-Tasking and Drive for Results/Openness to Change correlated closely at r=.583 and r=.559 respectively. The correlation of r=.366 between the HLI Achievement variable and Supervisor Rating Drive for Results was the highest correlation yielded in comparing the HLI assessment to the Supervisor Ratings.

Research Question Four – Data Analysis and Results

What other elements of the assessment instrument are significantly related to leadership performance ratings?

The final research question examined other points of interest regarding the two data sets. For example, a One-Way Analysis of Variance (ANOVA) was conducted to determine if the subjects' performance evaluation or HLI scores varied significantly depending upon level of management, years of service, or time in position (Salkind, 2004). Additional analysis was also conducted in this step including an examination of overall correlations between HLI and Supervisor Rating variables, Critical Thinking scores and the halo effect.

Table 23 shows the results of the ANOVA for the effects of Management Level, Years of Service, and Position Duration on Promotion Potential.

Table 23.

ANOVA for Management Level, Position and Employment Duration for Promo

		Sum of Squares	df	Mean Square	F	Sig.
Employment Duration	Between Groups	2.237	5	.447	.593	.705
	Within Groups	142.512	189	.754		
	Total	144.749	194			
Position Duration	Between Groups	10.719	5	2.144	1.771	.121
	Within Groups	228.799	189	1.211		
	Total	239.518	194			
Classification	Between Groups	3.178	5	.636	.712	.615
	Within Groups	156.258	175	.893		
	Total	159.436	180			

The large significance values of .121 to .705 in Table 23 show that the Potential for Promotion ratings did not differ significantly by Level of Management, Employment Duration or Position Duration.

Overall Correlations between all Study Variables

An examination of the relationships between all of the supervisor ratings and HLI assessment scores was also conducted, shown in Table 24.

		SR Drive Results	SR Conscientious	SR Cust Orient	SR Emotion Even	SR Innovativ	SR Multi-Tasking	SR Open Change	SR Self Develop	SR Self Confid	SR Potential Promo	SR Overall Perform	SR Ranking
HLI Achvmnt	Pearson	.366(**)	.219(**)	.229(**)	.167(*)	.231(**)	.204(**)	.253(**)	.237(**)	.170(*)	.241(**)	.248(**)	-.205(**)
	Sig. (2-tailed)	.000	.002	.001	.019	.001	.004	.000	.001	.017	.001	.000	.004
HLI Cust Orient	Pearson	.193(**)	.108	.309(**)	.251(**)	.086	.087	.188(**)	.074	.171(*)	.144	.218(**)	-.131
	Sig. (2-tailed)	.007	.133	.000	.000	.231	.229	.008	.303	.017	.044	.002	.067
HLI Conscientious	Pearson	.237(**)	.333(**)	.285(**)	.230(**)	.150(*)	.177(*)	.189(**)	.119	.100	.166(*)	.281(**)	-.282(**)
	Sig. (2-tailed)	.001	.000	.000	.001	.036	.013	.008	.099	.166	.021	.000	.000
HLI Emotion Even	Pearson	.073	-.056	.072	.219(**)	-.043	.032	.081	-.045	.135	.061	.079	-.002
	Sig. (2-tailed)	.310	.437	.314	.002	.549	.655	.258	.536	.060	.395	.272	.978
HLI Innovativ	Pearson	.223(**)	.086	.039	.102	.317(**)	.092	.205(**)	.112	.205(**)	.136	.172(*)	-.019
	Sig. (2-tailed)	.002	.230	.593	.158	.000	.201	.004	.118	.004	.058	.016	.793
HLI Multi-Tasking	Pearson	.171(*)	.217(**)	.049	.076	.136	.206(**)	.158(*)	.165(*)	.096	.181(*)	.182(*)	-.147(*)
	Sig. (2-tailed)	.017	.002	.496	.292	.059	.004	.027	.021	.183	.011	.011	.041
HLI Open to Change	Pearson	.195(**)	.192(**)	.093	.177(*)	.216(**)	.164(*)	.184(**)	.103	.161(*)	.199(**)	.208(**)	-.137
	Sig. (2-tailed)	.006	.007	.195	.013	.002	.022	.010	.154	.025	.005	.004	.057
HLI Self Confid	Pearson	.320(**)	.185(**)	.220(**)	.169(*)	.182(*)	.186(**)	.200(**)	.223(**)	.275(**)	.226(**)	.305(**)	-.177(*)
	Sig. (2-tailed)	.000	.010	.002	.018	.011	.009	.005	.002	.000	.001	.000	.013
HLI Self Develop	Pearson	.200(**)	.127	.140	.132	.158(*)	.095	.118	.218(**)	.084	.140	.172(*)	-.059
	Sig. (2-tailed)	.005	.076	.051	.065	.028	.186	.101	.002	.241	.051	.016	.413
Critical Think***	Pearson	.019	.153	.051	.065	.049	.109	.071	-.071	-.015	-.003	.049	-.097
	Sig. (2-tailed)	.827	.072	.549	.447	.565	.202	.403	.405	.862	.970	.566	.257

*** n=139 ** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Table 24 shows six correlations of more than r=.30, the strongest of which is between Supervisor Rating (SR) Drive for Results and the HLI Achievement score at r=.366. The second highest correlation was between SR Conscientiousness and the HLI Conscientiousness score (r=.333). Other correlations were SR Drive for Results – HLI Self Confidence (r=.320), SR Innovative – HLI Innovative (r=.317), SR Customer Focus - HLI Customer Focus (r=.309), and SR Overall Performance – HLI Self Confidence (r=.305). All of these correlations were significant at the $p<.01$ level. The HLI Self Confidence factor had two correlations above .30 with Supervisor Ratings (Drive for Results and Overall Performance), but not with the Supervisor Rating of Self-Confidence although they correlated at r=.275.

The significant correlations between the "mirrored" variables of supervisor ratings as compared to HLI scores are particularly notable for Achievement, Conscientiousness, Innovative, and Customer Focus for the two instruments. All of the mirrored HLI variables correlated with Supervisor Ratings at $p<.01$ significance, with the lowest correlation of r=.184 for Openness to Change.

Critical Thinking – A Second Look

Due to the extensive literature available which repeatedly attests to the predictive ability of general mental ability and cognitive reasoning in job performance and leadership selection (Lord, de Vader, & Alliger, 1986; Lubinski, 2004; Schmidt & Hunter, 2004; Vecchio, 1990), the need was identified to again review the critical thinking scores of the HLI. Because some research has shown a curvilinear relationship of critical thinking to other elements such as level of management (Smither & Reilly, 1987), curvilinear regression analysis was conducted between Critical Thinking and Potential for Promotion, as shown in Figure 7.

Figure 7.

Curve Fit Analysis of Critical Thinking Factor

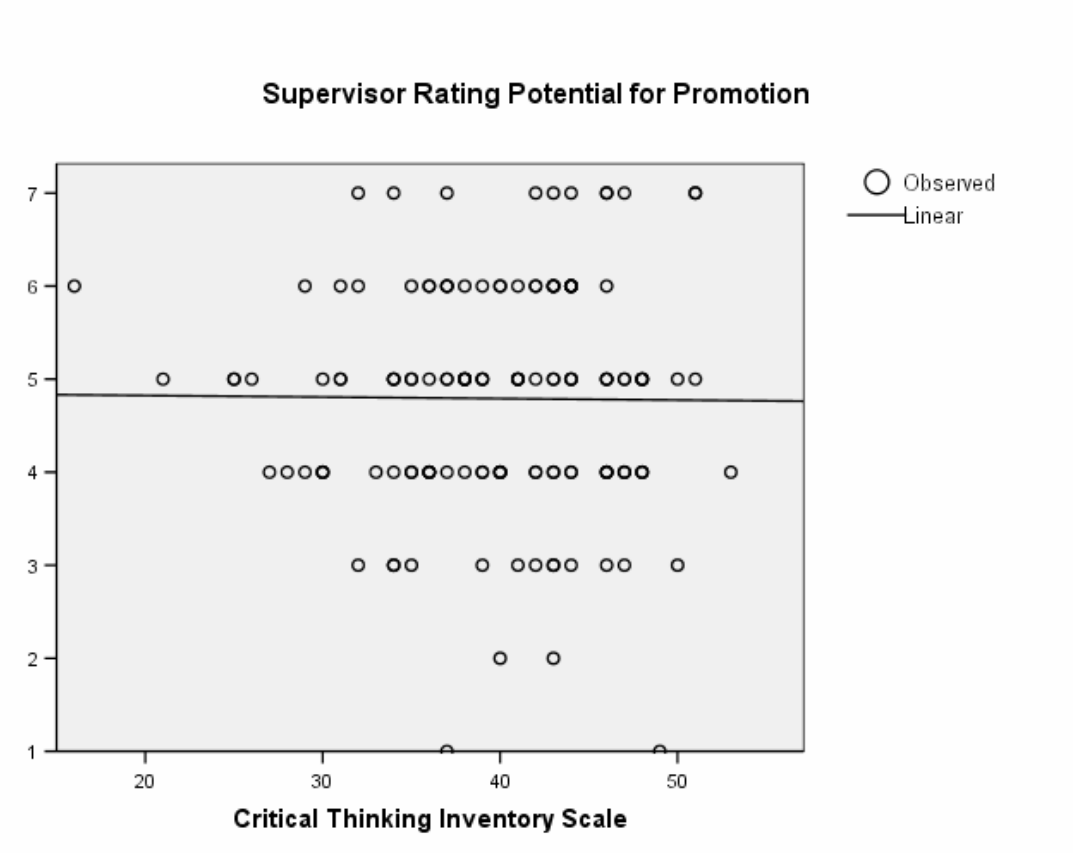


Figure 7 shows that there is no linear or curvilinear relationship between Promotion Potential and Critical Thinking.

Management Level was also stepwise regressed with Critical Thinking, as shown in Table 24.

Table 24.

Critical Thinking and Management Level Regression Analysis

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	39.439	1.718		22.955	.000
	Classification	.118	.485	.021	.244	.808

a Dependent Variable: Critical Thinking Inventory Scale

As Table 24 shows, there is no significant relationship between Critical Thinking and Management Level.

The Halo Effect

The HLI assessment creator anticipated the occurrence of halo effect in the performance appraisals. In an attempt to correct for this in the initial validity study, a cumulative performance score was derived for each subject by adding the Likert-scaled performance rating variables (J. E. Smith, 2006). This correction yielded higher scores (in The Kingwood Group's validity checks) in all of the HLI categories.

The halo effect can make it difficult for researchers to discern specific predictors of behavior because a supervisor who is under its influence is likely to give an employee specific ratings on the basis of an overall impression (positive or negative). The Potential for Promotion factor scores of the Supervisor Ratings had the lowest mean scores of all of the supervisory

ratings. While this does not guarantee the results of this study were not impacted by the halo effect, one can be relatively certain that supervisors were using a different standard. This research did not employ the cumulative performance score computed by the Kingwood Group because the target dependent variable was specifically Potential for Promotion.

Summary

The results of this study indicate considerable similarities in the constructs between the predictors yielded from the two instruments. The Performance Ratings yielded much higher correlations than the HLI assessment scores (as expected), but statistically significant relationships were identified in both. Critical Thinking surprisingly did not correlate with Promotion Potential in any meaningful way, although its presence or lack thereof affected the regression results. These results will be discussed in more length in Chapter Five. Additionally, conclusions and implications for future research will be addressed.

CHAPTER V - DISCUSSION

Chapter Five is organized in the following manner: a summary of the study, discussion and conclusions drawn from the results for each research question, contributions of the study, and implications for future research.

Summary of the Study

The purpose of this study was to identify predictors of leadership using a newly developed assessment for leadership selection within the healthcare industry by comparing assessment scores to supervisor rankings of the subjects. The study population consisted of 195 employees of 11 different hospitals. The participants completed the Healthcare Leadership Inventory; their immediate supervisors provided performance ratings on them. The researcher did not design any of the instruments used in this study.

The dependent variable of the study was the supervisor-provided factor of Promotion Potential. Stepwise multiple regression was the main analytical approach. The analysis yielded two predictors of leadership success from the HLI assessment (Achievement Orientation and Openness to Change) and five from the Supervisor Ratings (Multi-Tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation). The identified predictors from each instrument had construct symmetry, although they were not statistically duplicative. The predictors from Supervisor Ratings provided some insight into the implicit leadership theories shared by management personnel in the healthcare industry. Level of Management, Employment Duration, Position Duration, Gender, Race or Age of the subjects did not differ significantly when grouped by Potential for Promotion. The HLI assessment factors of Achievement, Conscientiousness, Innovative, and Customer Focus had significant correlations with their

counterparts from Supervisor Ratings. The Critical Thinking factor surprisingly did not significantly predict leadership potential or correlate with any of the other factors.

Research Question One – Discussion and Conclusions

What factors within the performance evaluation instrument are significantly related to promotion potential as rated by the subjects' supervisors?

The purpose of the first research question was to identify the strongest predictors of promotion based upon the supervisor ratings. Not only were these results compared to the results of Research Question Two, but they also provided insight into the implicit leadership theories of the supervisors who completed the ratings. The multiple regression analysis yielded five factors (Multi-Tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation) that accounted for 68.9% of the variation in Potential for Promotion. Therefore, these five factors were identified as the output of Research Question One.

These five factors are important on a number of fronts. First, they advanced to Research Question Three, where they were compared to the HLI assessment's predictive factors. Secondly, they offered insight into the implicit leadership theories of the supervisors who provided ratings of their employees. Although these supervisory raters were from 11 different hospitals, they systematically agreed that these dimensions were important for identifying leadership potential. Are they right? We do not know. One explanation of these results is that employees who demonstrate behaviors found to be consistent with Multi-tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation tend to be seen by their supervisors as having high promotion potential. Employees who are perceived to be high potentials (or "hi-pos") are likely to be provided with more opportunities to practice and improve their leadership skills, which can lead to additional promotions. The Pygmalion Effect, as discussed previously in this document and based upon Merton's (1957) self-fulfilling prophecy

theory, can help to explain how intuition and other less explicit forms of knowledge can affect perceived or real performance in individuals, which in turn affects decision-making about them.

As discussed in Chapter Three, the performance evaluation form completed by immediate supervisors of the subjects was not an artifact of their individual organizations and systems, but a creation of The Kingwood Group for the purpose of test validation. This circumstance could be either a blessing or curse. On the one hand, the literature states that people who complete performance ratings are more likely to be honest if they know the data will not be used to make decisions affecting pay, promotions, or disciplinary action (Coens & Jenkins, 2000; Murphy & Cleveland, 1995). This suggests that the performance data used in this study may be less contaminated than an artifact of an internal process might be. On the other hand, the supervisors were not familiar with the rating form they completed, nor did they receive any training on the constructs or their meanings. If they had been, would there have been stronger construct validity? Would the supervisor rating scores have more closely resembled the HLI scores? The answer is unknown; but the question would be an interesting one for future research.

Customer Orientation has been found to be an important predictor of success in other research. Taylor, Pajo, Cheung, and Stringfield (2004) used a Customer Service dimension in their research, and found that “only the customer focus dimension made a unique contribution to predicting job performance ratings” (761). Consistent with the literature, Customer Orientation showed predictive value in this study. One could conclude that either Customer Orientation was seen as important by the rating supervisors (or more as important as some other factors), or the characteristic is seen by supervisors (consciously or unconsciously) as relevant to Potential for Promotion.

Would a similar exercise with immediate supervisors of managers in another industry yield the same predictors of leadership? It is not known. However, based upon Edgar Schein's (1999) definition of culture as "the sum total of *all* the shared, taken-for-granted assumptions that a group has learned throughout its history" (p. 29), this process could potentially apply to entire industries just as it does to company culture. The healthcare industry is highly specialized and professionalized with advanced degrees. It is also more common to have females in leadership roles in the healthcare industry as compared to other industries due to the female dominated nursing profession. Although organizational or industrial culture is difficult to measure by all accounts, the idiosyncrasies of an industry could reveal themselves through a process similar to the one employed in Research Question One. This research step did not yield any "right" or "wrong" answers about leadership predictors – only that these five items are viewed (either consciously or subconsciously) as most predictive by the supervisors who provided the ratings. Whether right or wrong, the information gleaned from this step is rich in providing insight into the implicit leadership theories of the healthcare industry. Furthermore, the process used here could be replicated by any human resources practitioner with access to performance appraisal scores and statistical analysis software.

Research Question Two – Discussion and Conclusions

What factors within the HLI instrument are significantly related to promotion potential as rated by the subjects' supervisors?

The purpose of the second research question was to identify the strongest predictors of Promotion Potential based upon the HLI assessment scores. These results are useful as a comparison to the predictors identified in Research Question One, and they also provide another key insight. The correlations to Promotion Potential (Achievement Orientation and Openness to Change) were statistically significant. Together they accounted for over 10 percent of the

variation in the Promotion Potential variable. Therefore, the analysis identified Achievement Orientation and Openness to Change as the answer to Research Question Two. The relevance of two HLI assessment factors in the data analysis shows that the HLI factors do have some predictive ability for leadership potential. The multiple regression analysis yielded somewhat weak relationships by general statistical standards, but still significant by some field research standards (Barrick, Mount, & Gupta, 2003; Witt, Burke, Barrick, & Mount, 2002).

Would the correlations between the HLI scores and Promotion Potential have been stronger if the sample population and rating supervisors were all from the same company or if supervisors all had received training not only on how to complete performance appraisals, but also on the relevant constructs? This could be argued. However, most readers would agree that Potential for Promotion is a supervisory judgment that requires no training. On the other hand, among other things, organizational culture contributes to a supervisor's understanding and decision-making regarding promotion potential, and thus the associated ratings (Schein, 2004).

Heneman *et al.* suggested that relating assessment results to leadership development efforts is worthwhile. "For example, results may imply that training programs in selection decision-making would be desirable for staff and line managers, and the results may point toward desirable course content" (1980, p. 56).

Research Question Three – Discussion and Conclusions

How do the performance evaluation predictors compare to the HLI instrument predictors?

Research Question Three evaluated the relationships between the two sets of predictors both statistically and conceptually. All of the predictors (HLI-Achievement Orientation and HLI-Openness to Change, as well as Supervisor Ratings of Multi-Tasking, Drive for Results, Self Confidence, and Openness to Change) except one (Customer Orientation) are related to subsets

of the Big Five Personality Theory, which has been extensively evaluated and validated (Conway & Peneno, 1999; McCormack & Mellor, 2002; Morrison Jr., Abraham, & Dennis, 2004; Rubenzer, Faschingbauer, & Ones, 2000; Stricker & Rock, 1998; Tanoff & Barlow, 2002; Wielkiewicz, 2002). It is important to note that although the Kingwood Group provided initial information about how the constructs for the HLI were developed in relation to the Big Five Personality Theory, the researcher does not know the extent to which the HLI dimensions reflect the FFM. Therefore, only generalizations may be made here regarding the HLI's likeness to the FFM.

"Conscientiousness has emerged as the only general predictor of job performance, while other dimensions relate to more specific aspects of job performance" (John & Srivastava, 1999, p. 35). Conscientiousness certainly appears to be an important variable in selection; however it seems that it is a key *general* predictor for all job types. So, conscientiousness alone may be necessary but not sufficient for identifying good potential leaders.

One might suppose, given that the HLI Assessment and Performance Evaluation Form included the same dimensions (excepting Critical Thinking, Overall Performance, Ranking, and Potential for Promotion), that the mirrored factors from each instrument should have similarly varied with the dependent variable. This was not true for all of the variables. Although the reason is uncertain, it could be opined that the supervisors had not been trained on the meaning of the factors, which caused construct validity issues. However, construct validity problems are more pronounced with the HLI than with supervisor ratings, which are moderately inter-correlated, and yet diverge from potential for promotion.

One might also imagine that because the HLI is a self-rated instrument, a disparity could exist between how the immediate supervisors and employees saw the employees. While the HLI

does use a social desirability scale, the interpreted scales for social desirability were not available to the researcher. Therefore, the effects of social desirability on the subjects' HLI assessment scores cannot be presented at this time, nor can they be corrected for and compared to the supervisors' ratings to determine if this is the case.

In Chapter Four, a correlation analysis of the predictive factors yielded statistically significant ($p<.05$) correlations among all five factors. The correlations between the supervisor ratings and HLI scores ranged from r=.161 to r=.366. Descriptors of each of the key predictors were also compared to determine the overlap between supervisor rating predictors and the HLI predictors. Identical symmetry between the constructs was found for Goal Setting, Ambition, and Success. The other descriptor sets were very close: Accomplishment vs. Success, Adaptability vs. Juggling, and Encourages Others vs. Optimist/Enthusiastic. All of the predictors had conceptually similar counterparts on the other side of the equation. The fact that the predictors compared so closely is intriguing, and shows agreement between supervisors' implicit leadership theories and the HLI assessment instrument. This would have been an especially powerful finding if the HLI assessment scores had been responsible for more of the variation of Promotion Potential.

Research Question Four – Discussion and Conclusions

What other elements of the assessment instrument are significantly related to leadership performance ratings?

The sample scores did not appear to be biased in any way. Potential for Promotion distributed evenly among Level of Management, Years of Service, Position Duration, Gender, Race and Age of the subjects. The majority of the subjects were Caucasian (93.3 percent), female (74.4 percent), and 40 or more years of age (70.8 percent). Heneman *et al.* (1980) described a

tendency for females to be evaluated lower than males, “though the effects were not particularly strong” (p. 62), and it is good to see that this is not the case with the HLI.

Correlations among All Variables from Both Instruments

In the discussion regarding Research Question Three, it was noted that the mirrored factors from each instrument should have similarly varied with the dependent variable. While this was not true for all of the variables, significant correlations were found between the mirrored variables of the two instruments, particularly for Achievement, Conscientiousness, Innovative, and Customer Focus which all had correlations of over r=.30. Nunnelly and Bernstein (1994) asserted that correlations for individual personality factors rarely exceed r=.3 to r=.4 because measuring personality is such a complex endeavor. The correlations for these four factors all exceeded that threshold. In other words, the HLI dimensions did not correlate with the Dependent Variable as strongly as they did with their mirrored variables in the supervisor ratings. The significant relationships between the mirrored variables, while not the focus of this study, still provide positive evidence of the efficacy of the HLI instrument, particularly regarding the Achievement factor, which correlated significantly not only with the Dependent Variable, but also with its mirrored factor, Drive for Results, from the supervisors’ ratings. Clients of The Kingwood Group requested that it add a Customer Orientation facet to the instrument. In addition to the predictive value of Customer Orientation for Promotion Potential revealed in this study, the HLI assessment has also demonstrated the ability to measure Customer Focus in a way which is consistent with the supervisors’ view of Customer Focus.

The existence of correlations between the mirrored dimensions described above are indicative of convergent validity because “two independent methods of inferring an attribute [led] to similar ends” (Nunnally & Bernstein, 1994, p. 92). Discriminant validity also exists for

the HLI and the Supervisor Ratings to the extent that the variables from the two instruments did not tend to vary as strongly with non-mirrored factors as they did with their mirrored factors. For instance, HLI Innovative and SR Multi-Tasking had an r=.092 and p=.201, which is not at all a significant correlation. The presence of convergent and discriminant validities provides some evidence of construct validity for the HLI assessment.

Critical Thinking

The Critical Thinking factor surprisingly did not significantly predict leadership potential or correlate with any of the other factors. There are no certain explanations for this. There were no Critical Thinking scores for 56 of the subjects, so the subjects with these missing variables were not used in regression or correlation analysis which included the Critical Thinking scores. It is possible that these missing data are a symptom of some sort of sample bias, but there is no way of knowing this based upon the explanation provided by The Kingwood Group. Although validated against the Watson-Glaser Critical Thinking Assessment, the correlation was not high at r=.55 and the wording of the HLI's Critical Thinking items has been customized to the healthcare industry. There is a remote possibility that something was lost in the translation, but that should have been identified in the construct validity check if it were a problem.

Another possible explanation is that stress moderates the relationship between intelligence and leadership success, as suggested in Fiedler's Cognitive Resource Theory (Fiedler, 2001). Because this study did not examine stress levels as a component of the leadership experience, a possible explanation for the lack of relationship with any of the variables might have been overlooked. This would be an interesting area for future research, particularly within the healthcare arena where stress is a fact of daily life.

Because healthcare workers tend to be highly educated, it is also possible that no correlations were found because intelligence levels are more a threshold of entry into the management ranks in the healthcare arena than a differentiator among management ranks. Researchers who examine the critical thinking ability of healthcare-oriented managers as compared to other industry managers (particularly as the mean rates compare) will add to the body of knowledge on this topic. Whatever the reason for this study's failure to find correlations between Critical Thinking and other leadership traits, the Critical Thinking factor had no predictive validity with this particular sample.

The HLI Assessment Instrument

In his article "Who Should Lead a Healthcare Organization: MDs or MBAs?" Schultz *et al.* (2004) compared healthcare organizations led by "medically educated" and "managerially educated" executives and found that neither type of educational focus is more predictive of a good leader in healthcare than the other. However, the article also discussed a report of America's "Top 100 Hospitals" which sited the fact that "conspicuous among winners at every level were physician-led organizations" (106). If managerial education does not predict leadership success in an industry with so much specialization and constant environmental, competitive, and societal changes, what can? Having an assessment instrument to support practitioners' efforts is now more important than ever. With continued refinements as more populations take the test, additional predictive validity tests are conducted and the assessment is refined, so the HLI could be valuable in this regard.

There are a number of reasons why the HLI might become the assessment of choice for leaders in the healthcare industry. John and Srivastava (1999) note that a major advantage of using the FFM is that it is easily understandable in layperson's terms. This is certainly an

advantage with the HLI as well. Perhaps the most important attribute in its favor is that the HLI selection report includes a structured interview guide which provides probing questions to hiring managers to consider using based upon the assessment scores. The assessment has a "cheat scale" which provides detection for those who might try to "fake" the test. This cheat scale is extremely practical, providing decision-makers with one of three scores: "OK: This candidate responded to he assessment in a realistic manner," "CAUTION: This candidate may be attempting to make a favorable impression or may have an overly positive view of themselves," and "INVALID: These results are unusually positive. This candidate may be trying to 'cheat' the assessment." Such tools help to bridge the gap between practitioners and scholars by providing additional support with the administration and evaluation of the assessment, and by making its language and interpretation more accessible to non-scholars.

The HLI is a new assessment that has not yet undergone years and iterations of validity tests and adjustments ad have many tests which have been in existence for some time, and therefore might yield more statistically predictive results. Additionally, it is important to note that the test was not designed to be applied in the way in which it has in this research (using Potential for Promotion as the dependent variable); any results or statistics related to this research should not be interpreted as calling the test's efficacy into question.

The Big Five Personality Theory and the HLI

A.O. Bowden was one the first theorists to opine that personality and leadership were connected, stating "Indeed, the amount of personality attributed to an individual may not be unfairly estimated by the degree of influence he can exert upon others" (Bass & Stogdill, 1990, p. 12). Although many theories and theorists have since come and gone, one enduring framework which has broad acceptance for understanding personality has been the Big Five Personality

Theory (Barrick, Mount, & Gupta, 2003). The HLI Assessment's dimensions of Achievement and Openness to Change were developed based upon the Big Five constructs of Conscientiousness and Openness respectively. Conscientiousness has been called a general predictor of occupational success across a broad range of industries position levels (John & Srivastava, 1999, p. 35).

Sternberg's (2006) recent article highlighted three sets of studies that linked creativity lessons to students' school achievement. He emphasized that those students who either have the innate talent or who are trained to see technical problems as unbound by traditional frameworks have higher "Successful Intelligence" and the potential to outperform those who are less creative. Clearly, the HLI construct of Openness to Change is related to this concept and particularly to leadership. The fact that managers who were perceived by their supervisors to have more promotion potential achieved higher scores for Openness to Change shows that this mental flexibility is particularly important within the healthcare field.

Limitations

Limitations to the study include the use of one particular assessment instrument, and the examination of a specific industry (healthcare). In addition to the uniqueness of the healthcare industry already described in this study, the predominance of females in the sample population provides an additional reason that these results cannot be easily generalized beyond hospital populations. While females represent more than 50 percent of the United States workforce, they continue to be outnumbered by men in leadership positions in most industries (Marlowe, Schneider, & Nelson, 1996). Due to the prevalence of females in nursing positions in hospitals, the healthcare environment tends to be dissimilar to other industries in the gender makeup of its

leaders. The predominance of Caucasians in the sample also limits the ability to generalize the results into more diverse populations.

Because one cannot assess or estimate the effect that the degree of training the immediate supervisors of the subjects has had on performance evaluations or rating errors, the data regarding performance are subject to some construct validity concerns, as previously described. Additionally, due to the variability of the organizations within which the supervisors operate, it is difficult to know the extent to which the different work settings might account for variability in the ratings. However, a key to this research was to capture the implicit understandings of leadership potential that might be shared among supervisors in a healthcare setting; these performance evaluation data accomplished that quite effectively.

Regarding the implicit leadership theories of healthcare supervisors, this study has offered one approach to define commonalities in their perspectives. This study revealed that supervisors systematically agreed upon certain characteristics as being predictive of promotion potential, including Multi-tasking, Drive for Results, Self Confidence, Openness to Change, and Customer Orientation. However, the approach used here was extremely limited. To suggest that the supervisors' completion of an instrument in which they had no part in designing provides complete understanding of their implicit leadership theories would be an overstatement. Argyris (1978), Patton (1997), and Cooperrider et al. (2001) all describe methods of inquiry that are much more open to emergent themes. An improvement upon the current study would be to take a more qualitative approach to identifying the implicit leadership theories of supervisors, and comparing those to the HLI assessment results.

A goal in this study was to identify predictors of leadership that transcend organizational culture or individual rater bias in pursuit of more valid and reliable approaches to leadership

selection. The main idea was to help find more objective ways of leadership selection to augment existing selection processes. What is the meaning of the fact that immediate supervisor ratings were far more predictive of success than the HLI assessment results? Does this mean that there is no use for the HLI or assessments in general? It could mean a number of things, including the presence of the Pygmalion Effect, rater bias such as halo effect, Type I or Type II error, gender bias, construct validity issues, or a host of other factors for which researchers may or may not be able to account in the statistical analysis. Future research will continue to refine our understanding of the interactions of these elements.

Finally, the reader should also note that the test results are not a valid sole predictor of job performance. No test is a perfect measure of whether someone will be a good employee or leader. As discussed in Chapter One, hiring managers should place appropriate weighting on multiple data points for their hiring decisions.

Contributions of this Research

This study's contribution to the work of scholars and practitioners on leadership selection is threefold. First, some factors of the FFM were confirmed as predictors of leadership for the healthcare industry. Secondly, the data analysis and results conducted in this study help to increase understanding of a new instrument in the healthcare industry, and also help to refine the measurement of predictors going forward. Finally, this research modeled a process which could be used by other practitioners for assessing predictive validity of assessments used within their own organizations or industries.

Implications for Future Research

As noted elsewhere in this chapter, it would be interesting to see if there were an impact of managers' scoring of subordinates if these managers were trained in leadership and selection.

Further research is suggested on the effects of social desirability, halo effect, and rater bias. Because stress is an important component in the lives of healthcare leaders due to the nature of the healthcare industry, further exploration into the effects of stress as a moderator between intelligence and leadership performance would be useful. Perhaps most important is the need for continuation of predictive research that uses reliable measures of effective leadership across organizational and industrial cultures.

Assumptions

The intent of this research was to provide better understanding about predictors of leadership potential so that practitioners would have more objective selection tools for their organizations. In attempting to do so, however, the researcher does not claim that having assessments which yield more predictive results should reduce the amount of intuition that practitioners currently apply in their selection decisions. The goal is not to create a process that *removes* subjectivity from the leadership selection equation. Hiring decisions are ultimately value-based; they are inherently subject to personal judgment which relies on the use of tacit knowledge as well as explicit knowledge. Schön (1983) documented a clear distinction between "technical rationality" and "reflection in action." Just as exclusively positivistic research methods have left scientists wanting for more qualitative approaches to inquiry, as well as broader acceptance of many ways of knowing, it is important to recognize that hiring decisions are not based solely upon technical questions with standardized inputs. Selection decisions, like research questions, are based upon a mixture of data points, judgments, and values, some of which may very well conflict. No selection tool can completely objectify the hiring decision, nor should decision makers desire such an implement. Any factor which feeds the hiring decision,

including an assessment instrument, should be but one element of a widely understood array of "evidence" on the matter. The identification and use of each element should be accomplished through a reflective and deliberate process, such as that which has been attempted in this dissertation.

It also must be made clear that, while there are various levels of sophistication among HR professionals, this study was not undertaken to suggest that HR professionals are not capable of understanding and employing scholarly research. It should also be recognized that predicting human factors such as performance can be a complicated business. There are also times in which employing an expert in psychometric techniques would be wise.

Summary

It is clear that there is a fundamental gap between scholars and practitioners in shared knowledge about leadership and leadership selection. While the differences are great, there is also a good deal of common ground from which to begin to build a bridge between the two worlds. This study offered insight into the implicit leadership theories that are shared among supervisors within the healthcare industry, as revealed in analysis of performance evaluations. Once this implicit information was made explicit through the data analysis, it was compared to external measures (an assessment in this case). The absence of some factors and presence of others has provoked some dissonance, and has also somewhat confirmed the existing implicit theories of leadership in this industry. While the actual results of this research process might hold potential for improving leadership selection tools within the healthcare industry, the larger goal is to facilitate a more informed discussion about the utility of using various measures in selection.

REFERENCES

Ahlburg, D. A. (1992). Predicting the job performance of managers: What do the experts know? *International Journal of Forecasting, 7*(4), 467-473.

Alban-Metcalfe, R. J., & Alimo-Metcalfe, B. (2000a). An analysis of the convergent and discriminant validity of the Transformational Leadership Questionnaire. *International Journal of Selection and Assessment, 8*(3), 158-175.

Alban-Metcalfe, R. J., & Alimo-Metcalfe, B. (2000b). The transformational leadership questionnaire (TLQ-LGV): A convergent and discriminant validation study. *Leadership & Organization Development Journal, 21*(6), 280.

Alban Metcalfe, R. J., & Alimo Metcalfe, B. (2000). An analysis of the convergent and discriminant validity of the Transformational Leadership Questionnaire. *International Journal of Selection and Assessment, 8*(3), 158-175.

Alimo-Metcalfe, B., & Alban-Metcalfe, R. J. (2001). The development of a new transformational leadership questionnaire. *Journal of Occupational and Organizational Psychology, 74*(1), 1-27.

Allen, S. J. (2006). *An exploration of theories of action in leadership development: A case study*. Unpublished Dissertation, Antioch University, Yellow Springs, OH.

Altman, D. (2006). Quality leadership: A conversation with Dennis. R. Barry. *Leadership in Action, 26*(1), 14-18.

The American Heritage Dictionary. (Fourth ed.)(2001). Fourth ed.). New York: Dell Publishing.

American Management Association survey on workplace testing. (1999). *Management Review, 88*(7), 44-47.

Argyris, C., & Schon, D. (1978). *Theory in practice: Increasing professional effectiveness*. San Fransisco: Jossey-Bass.

Aristotle, & Burnet, J. (1973). *The ethics of Aristotle*. New York: Arno Press.

Ashworth, S. D., & Hazer, J. T. (1986). Scale content, respondent classification, referent stereotyping, and stability in four LPC measures. *Educational and Psychological Measurement, 46*(1), 107-123.

Atwater, L. E., & Yammarino, F. J. (1993). Personal attributes as predictors of superiors' and subordinates' perceptions of military academy leadership. *Human Relations, 46*(5), 645-668.

Avolio, B. J., Bass, B. M., & Jung, D. I. (1999). Re-examining the components of transformational and transactional leadership using the Multifactor Leadership Questionnaire. *Journal of Occupational and Organizational Psychology, 72*(4), 441-462.

Avolio, B. J., Howell, J. M., & Sosik, J. J. (1999). A funny thing happened on the way to the bottom line: Humor as a moderator of leadership style effects. *Academy of Management Journal, 42*(2), 219-227.

Ayman, R., & Chemers, M. M. (1991). The effect of leadership match on subordinate satisfaction in Mexican organisations: Some moderating influences of self-monitoring. *Applied Psychology: An International Review, 40*(3), 299-314.

Babcock, P. (2003). Spotting lies. *HR Magazine, 48*(10), 46-52.

Bakker, A. B., Van Der Zee, K. I., Lewig, K. A., & Dollard, M. F. (2006). The relationship between the big five personality factors and burnout: A study among volunteer counselors. *Journal of Social Psychology, 146*(1), 31-50.

Baril, G. L., Ayman, R., & Palmiter, D. J. (1994). Measuring leader behavior: Moderators of discrepant self and subordinate descriptions. *Journal of Applied Social Psychology, 24*(1), 82-94.

Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology, 56*(1), 45-.

Bass, B. M. (1981). *Bass & Stogdill's handbook of leadership: theory, research, and managerial applications* (3rd ed.). New York: The Free Press.

Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York, London: Free Press; Collier Macmillan.

Bass, B. M., Avolio, B. J., & Goodheim, L. (1987). Biography and the assessment of transformational leadership at the world-class level. *Journal of Management, 13*(1), 7-19.

Bass, B. M., & Stogdill, R. M. (1990). *Bass & Stogdill's handbook of leadership: theory, research, and managerial applications* (3rd ed.). New York, London: Free Press; Collier Macmillan.

Bennis, W. G., & Nanus, B. (1985). *Leaders: The strategies for taking charge* (1st ed.). New York: Harper & Row.

Bhal, K. T., & Ansari, M. A. (1996). Measuring quality of interaction between leaders and members. *Journal of Applied Social Psychology, 26*(11), 945-972.

Blake, R. R., & Mouton, J. S. (1964). *The managerial grid: Key orientations for achieving production through people*. Houston, Tex.,: Gulf Pub. Co.

Bliese, P. D., Halverson, R. R., & Schriesheim, C. A. (2002). Benchmarking multilevel methods in leadership: The articles, the model, and the data set. *Leadership Quarterly, 13*(1), 3-14.

Boulden, G. B., Terence; Bruce, Andy; Eaton, John; Holden, Robert; Johnson, Roy; Langdon, Ken; Osborne, Christina; Renshaw, Ben; Seymour, John; Shervington, Martin; Tee, Rebecca; Ali, Moi (Editor). (2002). *Successful manager's handbook (Financial Times (DK)) (Hardcover)*. New York: DK Publishing, Inc.

Braun, H. I., Jackson, D. N., Wiley, D. E., & Messick, S. (2002). *The role of constructs in psychological and educational measurement*. Mahwah, N.J.: L. Erlbaum Associates.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the Internet. *European Journal of Psychological Assessment, 21*(2), 115-127.

Buckley, M. R., Ferris, G. R., Bernardin, H. J., & Harvey, M. G. (1998). The disconnect between the science and practice of management. *Business Horizons*.

Burns, J. M. (1978). *Leadership* (1st ed.). New York: Harper & Row.

Butler, J. K. (1993). Assessing the situations of the LEAD. *Organization Development Journal, 11*(4), 33-42.

Bycio, P., Hackett, R. D., & Allen, J. S. (1995). Further assessments of Bass's (1985) conceptualization of transactional and transformational leadership. *Journal of Applied Psychology, 80*(4), 468-478.

Capra, F. (1996). *The web of life: A new scientific understanding of living systems* (1st Anchor Books ed.). New York: Anchor Books.

Carless, S. A. (2001). Assessing the discriminant validity of the Leadership Practices Inventory. *Journal of Occupational and Organizational Psychology, 74*(2), 233-239.

Cartwright, D., & Zander, A. F. (1953). *Group dynamics, research and theory*. Evanston, Illinois: Row.

Cascio, W. F. (1989). *Managing human resources: Productivity, quality of work life, profits* (2nd ed.). United States of America: McGraw-Hill.

Chemers, M. M., Watson, C. B., & May, S. T. (2000). Dispositional affect and leadership effectiveness: A comparison of self-esteem, optimism, and efficacy. *Personality and Social Psychology Bulletin, 26*(3), 267-277.

Code of Federal Regulations, 41 CFR 60 - 3.4 - Information on impact. (2006). Retrieved March 12, 2006, 2006, from http://www.dol.gov/dol/allcfr/Title_41/Part_60-3/41CFR60-3.4.htm

Coens, T., & Jenkins, M. (2000). *Abolishing performance appraisals: Why they backfire and what to do instead*. San Fransisco, CA: Berrett-Koehler Publishers, Inc.

Confucius, & Slingerland, E. G. (2003). *Confucius analects: With selection from traditonal commentaries*. Indianapolis, Indiana: Hackett Publishing Company.

Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology, 13*(4), 485-506.

Cooperrider, D. L., Sorensen Jr., P. F., Yaeger, T. F., & Whitney, D. (2001). *Appreciative inquiry*. Champaign, Illinois: Stipes Publishing, LLC.

Davidson, O. B., & Eden, D. (2000). Remedial self-fulfilling prophecy: Two field experiements to prevent Golem effects among disadvantaged women. *Journal of Applied Psychology, 85*(3), 386-398.

de Vries, R. E., Roe, R. A., & Taillieu, T. C. B. (1998). Need for supervision: Its impact on leadership effectiveness. *Journal of Applied Behavioral Science, 34*(4), 486-501.

Decker, W. H., & Rotundo, D. M. (1999). Use of humor at work: Predictors and implications. *Psychological Reports, 83*(3, Part 1), 961-968.

Denzin, N. K., & Lincoln, Y. S. (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, Calif.: Sage Publications.

Dessler, G. (2003). *Human resource management* (9th ed.). Upper Saddle River, New Jersey: Pearson Education.

Eden, D., & Shani, A. B. (1982). Pygmalion goes to boot camp: Expectancy, leadership, and trainee performance. *Journal of Applied Psychology, 67*(2), 194-199.

Edwards, A. L. (1957). *The social desirability varibale in personality assessment and research*. New York: Dryden Press.

Edwards, J. E., Rode, L. G., & Ayman, R. (1989). The construct validity of scales from four leadership questionnaires. *Journal of General Psychology, 116*(2), 171-181.

Eiser, B. J. A., Eiser, A., R., & Parmer, M. A. (2006). Power of persuasion: Influence tactics for health care leaders. *Leadership in Action, 26*(1), 3-7.

Farh, J. L., Podsakoff, P. M., & Cheng, B. S. (1987). Culture-free leadership effectiveness versus moderators of. *Journal of International Business Studies, 18*(3), 43.

Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.

Fiedler, F. E. (2001). When IQ + experience = performance. *Leadership & Organizational Development Journal, 22*(3), 132.

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage.

Fields, D. L., & Herold, D. M. (1997). Using the Leadership Practices Inventory to measure transformational and transactional leadership. *Educational and Psychological Measurement, 57*(4), 569-579.

Folger, R., & Cropanzano, R. (1998). *Organizational justice and human resource management*. Thosand Oaks, CA: Sage.

Follert, V. (1983). Supervisors' power: An exchange model of leadership. *Psychological Reports, 52*(3), 740.

Ford, E. W., Duncan, W. J., Bedeian, A. G., Ginter, P. M., Rousculp, M. D., & Adams, A. M. (2005). Mitigating risks: visible hands, inevitable disasters, and soft variables: Management research that matters to managers. *Academy of Management Executive, 19*(4), 24-38.

Fox, W. M., Hill, W. A., & Guertin, W. H. (1973). Dimensional analysis of the Least Preferred Co-Worker scales. *Journal of Applied Psychology, 57*(2), 192-194.

Fuller, J. B., Patterson, C. E. P., Hester, K., & Stringer, D. Y. (1996). A quantitative review of research on charismatic leadership. *Psychological Reports, 78*(1), 271-287.

Garman, A. N., & Tran, L. (2006). Knowledge of the healthcare environment. *Journal of Healthcare Management, 51*(3), 152-155.

Garman, A. N., Tyler, J. L., Darnall, J. S., & Lerner, W. (2004). Development and validation of a 360-degree-feedback instrument for healthcare administrators. *Journal of Healthcare Management, 49*(5), 307-322.

George, D., & Mallery, P. (2006). *SPSS for Windows step by step: A simple guide and reference* (Sixth ed.). Boston, MA: Pearson Education, Inc.

Gilliland, S. W. (1993). The Perceived fairness of selection systems; An organizational justice perspective. *Academy of Management Review, 18*(4), 694-734.

Gioia, D. A., & Sims, H. P. (1985). On avoiding the influence of implicit leadership theories in leader behavior descriptions. *Educational and Psychological Measurement, 45*(2), 217-232.

Gladwell, M. (2005). *Blink: The power of thinking without thinking* (1st ed.). New York: Little, Brown and Co.

Glazer, N. (1983). *Ethnic dilemmas, 1964-1982*. Cambridge, Mass.: Harvard University Press.

Glisson, C., & James, L. R. (2002). The cross-level effects of culture and climate in human service teams. *Journal of Organizational Behavior, 23*, 767-794.

Goffee, R., & Jones, G. (1998). *The character of a corporation: How your company's culture can make or break your business* (1st ed.). New York: HarperBusiness.

Goldberg, L. R. (1990). An alternative 'description of personality': The big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Goodwin, V. L., Wofford, J. C., & Whittington, J. L. (2001). A theoretical and empirical extension of the transformational leadership construct. *Journal of Organizational Behavior, 22*(7), 759-774.

Green, S. G., Nebeker, D. M., & Boni, M. A. (1976). Personality and situational effects on leader behavior. *Academy of Management Journal, 19*(2), 184.

Grugulus, I. (2002). Nothing serious? Candidates' use of humor in management training. *Human Relations, 55*(4), 387-406.

Hammonds, K. H. (2005). Why we hate HR. *Fast Company*(97).

Harland, L. K. (2003). Using personality tests in leadership development: Test format effects and the mitigating impact of explanations of feedback. *Human Resource Development Quarterly, 14*(3), 285-301.

Head, T. C. (1991). An examination of the priming effect as it pertains to the Leader Behavior Description Questionnaire Form XII. *Psychological Reports, 68*(2), 515-520.

Heifetz, R. A. (1994). *Leadership without easy answers*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Heilman, M. E. (1997). Sex discrimination and the affirmative action remedy: The role of sex stereotypes. *Journal of Business Ethics, 16*(9), 877.

Heneman III, H. G., Hamstra, B. W., & Brown, S. H. (1980). Role of valid test score information and other factors in selection decision-making. *Journal of Management, 6*(1), 55-64.

Higgs, M., & Aitken, P. (2003). An exploration of the relationship between emotional intelligence and leadership potential. *Journal of Managerial Psychology, 18*(8), 814-823.

Houghton, J. D., & Neck, C. P. (2002). The revised self-leadership questionnaire: Testing a hierarchical factor structure for self-leadership. *Journal of Managerial Psychology, 17*(8), 672-691.

House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management, 23*(3), 409-473.

How useful are normality tests? (2006). Retrieved 7/28/2006, 2006, from http://www.graphpad.com/library/BiostatsSpecial/article_197.htm

Huffcutt, A. I., & Arthur, W., Jr. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*(2), 184-190.

Hughes, M. M., Patterson, L. B., & Terrell, J. B. (2005). *Emotional intelligence in action: Training and coaching activities for leaders and managers*. San Francisco: Pfeiffer.

Ilgen, D. R., & O'Brien, G. (1974). Leader-member relations in small groups. *Organizational Behavior & Human Performance, 12*(3), 335-350.

Irvine, D., Leatt, P., Evans, M. G., & Baker, R. G. (1999). Measurement of staff empowerment within health service organizations. *Journal of Nursing Measurement, 7*(1), 79-96.

Jaeger, R. M. (1993). *Statistics: A spectator sport* (2nd ed.). Newbury Park, CA: Sage Publications, Inc.

John, O. P., & Srivastava, S. (1999). The big-five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality Theory and Research, 2*, 102-138.

Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology, 89*(3), 428-441.

Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology, 89*(3), 542-552.

Kellerman, B. (2004). *Bad leadership: What it is, how it happens, why it matters*. Boston: Harvard Business School Press.

Kennedy, K, J., & Jr. (1982). Middle LPC leaders and the contingency model of leadership effectiveness. *Organizational Behavior and Human Performance, 30*(1), 1.

Kennedy, J. K., & Gallo, D. D. (1986). Test re-test properties of the Least Preferred Co-Worker (LPC) score. *Journal of Psychology: Interdisciplinary and Applied, 120*(6), 607-612.

Kennedy, J. K., Houston, J. M., Korsgaard, M. A., & Gallo, D. D. (1987). Construct space of the Least Preferred Co-worker (LPC) Scale. *Educational and Psychological Measurement, 47*(3), 807-814.

Kerr, S., & Jermier, J. M. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance, 22*,, 375–403.

Kierein, N. M. (2000). Pygmalion in work organizations: A meta-analysis. *Journal of Organizational Behavior, 21*(8), 913.
Klakovich, M. (1995). Development and psychometric evaluation of the Reciprocal Empowerment Scale. *Journal of Nursing Measurement, 3*(2), 127-143.
Kline, P. (2000). *A psychometrics primer*. London: Free Association Books.
Knight, W. H., & Holen, M. C. (1985). Leadership and the perceived effectiveness of department chairpersons. *Journal of Higher Education, 56*(6), 677-690.
Koene, B. A. S., Vogelaar, A. L. W., & Soeters, J. L. (2002). Leadership effects on organizational climate and financial performance: Local leadership effect in chain organizations. *The Leadership Quarterly, 13*, 193-215.
Komarraju, M., & Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Personality and Individual Differences, 39*, 557-567.
Kouzes, J. M., & Posner, B. Z. (1993). *Leadership practices inventory: Trainer's manual* (1993 expanded ed.). Amsterdam, San Diego: Pfeiffer.
Kravitz, S. M., & Schubert, S. D. (2000). *Emotional intelligence works: Developing "people smart" strategies*. Menlo Park, CA: Crisp Learning.
Lam, & K, S. S. (1998). An assessment of the reliability and validity of the leadership practices inventory in Hong Kong. *International Journal of Management, 15*(1), 57.
Leung, S. L., & Bozionelos, N. (2003). Five-factor model traits and the prototypical image of the effective leader in the Confucian culture. *Employee Relations, 26*(1), 62-71.
Lievens, F., Highhouse, S., & Corte, W. D. (2005). The importance of traits and abilities in supervisors' hirability decisions as a function of method of assessment. *Journal of Occupational and Organizational Psychology, 78*, 453-470.
Lord, R. G., de Vader, C. L., & Alliger, G. M. (1986). A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology, 71*(3), 402-410.
Loveland, J. M., Gibson, L. W., Lounsbury, J. W., & Huffstetler, B. C. (2005). Broad and narrow personality traits in relation to the job performance of camp counselors. *Child and Youth Care Forum, 34*(3), 241-255.
LPI Online - A Leadership Challenge Resource. (2002). Retrieved November 15, 2005, 2005, from http://www.lpionline.com/lpi/helpInfo/methodology.jsp
Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's "'General Intelligence,' objectively determined and Measured". *Journal of Personality and Social Psychology, 86*(1), 96-111.
Lueder, D. C. (1985). Don't be misled by LEAD. *Journal of Applied Behavioral Science, 21*(2), 143-151.
Lynn, A. B., & Lynn, A. B. (2002). *The emotional intelligence activity book: 50 activities for developing EQ at work*. New York: AMACOM.
MacCann, C., Matthews, G., Zeidner, M., & Roberts, R. (2003). Psychological assessment of emotional intelligence: A review of self-report and performance-based testing. *International Journal of Organizational Analysis, 11*(3), 247-268.
Machiavelli, N., & Bull, G. A. (2005). *The prince*. New York: Penguin Books.
Mapes, K. (2000). *Stop! think! choose!: Building emotional intelligence in young people*. Tucson, Ariz.: Zephyr Press.
March, J. G., Simon, H. A., & Guetzkow, H. S. (1993). *Organizations* (2nd ed.). Cambridge, Mass., USA: Blackwell.

Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology, 81*(1), 11-21.

McCall, M. W., Lombardo, M. M., & Morrison, A. M. (1988). *The lessons of experience: How successful executives develop on the job*. Lexington, MA: Lexington Books.

McClelland, D. C., & Boyatzis, R. E. (1982). Leadership motive pattern and long-term success in management. *Journal of Applied Psychology, 67*(6), 737-743.

McCormack, L., & Mellor, D. (2002). The role of personality in leadership: An application of the Five-Factor Model in the Australian military. *Military Psychology, 14*(3), 179-197.

McCrae, R. R., & Costa Jr., P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509-516.

McGilton, K. S. (2003). Development and psychometric evaluation of supportive leadership scales. *CJNR: Canadian Journal of Nursing Research, 35*(4), 72-86.

Merlevede, P. E., Bridoux, D., & Vandamme, R. (2003). *7 steps to emotional intelligence*. Carmarthen, Wales ; Williston, VT: Crown House Pub.

Merton, R. K. (1957). *Social theory and social structure* (Rev. and enl. ed.). Glencoe, Ill.: Free Press.

Morrison Jr., J. D., Abraham, J. D., & Dennis, G. B. (2004). Relationship between the Performance Perspectives Inventory's leadership scales and job performance of quick service restaurant managers. *Applied H.R.M.Research, 9*(1), 31-34.

Mullen, B., Symons, C., Hu, L. t., & Salas, E. (1989). Group size, leadership behavior, and subordinate satisfaction. *Journal of General Psychology, 116*(2), 155-170.

Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal : social, organizational, and goal-based perspectives*. Thousand Oaks, Calif.: Sage Publications.

Newcombe, M. J., & Ashkanasy, N. M. (2002). The role of affect and affective congruence in perceptions of leaders: An experimental study. *The Leadership Quarterly, 13*, 601-614.

Northouse, P. G. (1997). *Leadership: Theory and practice*. Thousand Oaks, CA: Sage Publications.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

O'Reilly, C. A., & Pfeffer, J. (2000). *Hidden value: How great companies achieve extraordinary results with ordinary people*. Boston, MA: Harvard Business School Press.

O'Reilly III, C. A., & Pfeffer, J. (1987). Hospital demography and turnover among nurses. *Industrial Relations, 26*(2), 158-174.

O'Reilly III, C. A., & Pfeffer, J. (2000). Cisco Systems: Acquiring and retaining talent in hypercompetitive markets. *HR. Human Resource Planning, 23*(3), 38-53.

O'Reilly III, C. A., & Pfeffer, J. (2000). Unlocking the hidden value in organizations. *Employment Relations Today, 27*(2), 63.

O'Toole, J. (1996). *Leading change: The argument for values-based leadership*. New York, NY: Jossey-Bass.

Oh, M.-Y., & Ramaprasad, J. (2003). Halo effect: Conceptual definition and empirical exploration with regard to South Korean subsidiaries of US and Japanese multinational companies. *Journal of Communication Management, 7*(4), 317-330.

Orpen, C., & Hall, C. (1994). Testing the "substitutes" model of managerial leadership: An examination of moderators of the relationships between leader reward and punishment behaviours and employee job satisfaction. *Studia Psychologica, 36*(1), 65-68.

Patton, M. Q. (1997). *Utilization focused evaluation: The new century text*. Thousand Oaks, CA: Sage.

Paul, A. M. (2004). *The cult of personality: How personality tests are leading us to miseducate our children, mismanage our companies, and misunderstand ourselves*. New York: Free Press.

Peters, L. H., Hartke, D. D., & Pohlmann, J. T. (1985). Fiedler's contingency theory of leadership: An application of the meta-analysis procedures of Schmidt and Hunter. *Psychological Bulletin, 97*(2), 274-285.

Philbrick, J. H., Bart, B. D., Sparks, M. R., & Hass, M. E. (1999). Pre-employment screening: A decade of change. *American Business Review, 17*(2), 75-85.

Pitner, N. J. (1988). Leadership substitutes: Their factorial validity in educational organizations. *Educational and Psychological Measurement, 48*(2), 307-315.

Plato, & Grube, G. M. A. (1974). *The Republic*. Indianapolis: Hackett Pub. Co.

Podsakoff, P. M., & MacKenzie, S. B. (1994). An examination of the psychometric properties and nomological validity of some revised and reduced substitutes for leadership scales. *Journal of Applied Psychology, 79*(5), 702-713.

Podsakoff, P. M., MacKenzie, S. B., & Bommer, W. H. (1996). Meta-analysis of the relationships between Kerr and Jermier's substitutes for leadership and employee job attitudes, role perceptions, and performance. *Journal of Applied Psychology, 81*(4), 380-399.

Podsakoff, P. M., Niehoff, B. P., MacKenzie, S. B., & Williams, M. L. (1993). Do substitutes for leadership really substitute for leadership? An empirical examination of Kerr and Jermier's situational leadership model. *Organizational Behavior and Human Decision Processes, 54*(1), 1-44.

Podsakoff, P. M., Todor, W. D., Grover, R. A., & Huber, V. L. (1984). Situational moderators of leadership reward and punishment behaviors: Fact or fiction? *Organizational Behavior & Human Performance, 34*, 21-63.

Posner, B. Z. (2004). A Leadership development instrument for students: Updated. *Journal of College Student Development, 45*(4), 443-456.

Posner, B. Z., & Brodsky, B. (1992). A leadership development instrument for college students. *Journal of College Student Development, 33*(3), 231-237.

Principles for the validation and use of personnel selection procedures. (2003).). Bowling Green, OH: Society for Industrial Organizational Psychology.

Rheem, H. (Cartographer). (1995). *Effective leadership: The Pygmalion effect*

Rogers, R. W., & Smith, A. B. (2004). Spotting executive potential and future senior leaders. *Employment Relations Today, 31*(1), 51-60.

Rose, D. S., & Fiore, K. E. (1999). Practical considerations and alternative research methods for evaluating HR programs. *Journal of Business and Psychology, 14*(2), 235-251.

Rosenthal, R. (1997, August 16, 1997). *Interpersonal expectancy effects: A forty year perspective.* Paper presented at the American Psychological Association Convension, Chicago, IL.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development.* (Vol. xi,). New York, NY, US: Holt, Rinehart & Winston, Inc.

Roskin, R., & Margerison, C. (1983). The effectiveness of some measures of managerial effectiveness. *Human Relations, 36*(10), 865-882.

Rost, J. C. (1991). *Leadership for the twenty-first century*. New York: Praeger.
Rubenzer, S. J., Faschingbauer, T. R., & Ones, D. S. (2000). Assessing the U.S. presidents using the revised NEO Personality Inventory. *Assessment, 7*(4), 403-420.
Rust, J., & Golombok, S. (2004). *Modern psychometrics: The science of psychological assessment* (Second edition ed.). New York, NY: Routledge.
Ryback, D. (1998). *Putting emotional intelligence to work: Successful leadership is more than IQ*. Boston: Butterworth-Heinemann.
Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*(2), 149-174.
Sala, F., & Dwight, S. A. (2002). Predicting executive performance with multirater surveys: Whom you ask makes a difference. *Consulting Psychology Journal: Practice and Research, 54*(3), 166-172.
Sala, F., Urch Druskat, V., & Mount, G. (2006). *Linking emotional intelligence and performance at work: Current research evidence with individuals and groups*. Mahwah, N.J.: Lawrence Erlbaum Associates.
Salkind, N. J. (2004). *Statistics for people who (think they) hate statistics* (2nd ed.). Thousand Oaks, Calif.: Sage Publications.
Schein, E. H. (1978). *Career dynamics: Matching individual and organizational needs*. Reading, MA: Addison-Wesley Pub. Co.
Schein, E. H. (1992). *Organizational culture and leadership* (2nd ed.). San Francisco: Jossey-Bass.
Schein, E. H. (1999). *The corporate culture survival guide: Sense and nonsense about culture change* (1st ed.). San Francisco, Calif.: Jossey-Bass.
Schein, E. H. (2004). *Organizational culture and leadership* (3rd ed.). San Francisco: Jossey-Bass.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.
Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work; Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162-173.
Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
Schriesheim, A, C., DeNisi, & S, A. (1981). Task dimensions as moderators of the effects of instrumental leadership: A two-sample replicated test of Path-Goal Leadership Theory. *The Journal of Applied Psychology, 66*(5), 589.
Schriesheim, C. A., Tepper, B. J., & Tetrault, L. A. (1994). Least preferred co-worker score, situational control, and leadership effectiveness: A meta-analysis of contingency model performance predictions. *Journal of Applied Psychology, 79*(4), 561-573.
Schultz, F. C. (2004). Who should lead a healthcare organization: MDs or MBAs? *Journal of Healthcare Management, 49*(2), 103-117.
Schultz, F. C., Pal, S., & Swan, D. A. (2004). Who should lead a healthcare organization: MDs or MBAs? *Journal of Healthcare Management, 49*(2), 103-117.
Seltzer, J., & Bass, B. M. (1990). Transformational leadership: Beyond initiation and consideration. *Journal of Management, 16*(4), 693-703.

Shamir, B., & Zakay, E. (1998). Correlates of charismatic leader behavior in military units: Subordinates' attitudes, unit.. *Academy of Management Journal, 41*(4), 387.

Shiflett, & Samuel. (1981). Is there a problem with the LPC score in leader match? *Personnel Psychology, 34*(4), 765.

Smith, C., Cowan, C., Heffler, S., & Catlin, A. (2006). National health spending in 2004: Recent slowdown led by prescription drug spending. *Health Affairs, 25*(1), 186-196.

Smith, J. E. (2006). *Technical manual for the healthcare leadership inventory (in press).* Hartland, MI

Smither, J. W., & Reilly, R. R. (1987). Ture intercorrelation among job components, time delay in rating, and rater intelligence as determinants of accuracy in performance ratings. *Organizational Behavior & Human Decision Processes, 40*(3), 369-392.

Spangenberg, H. H., & Theron, C. C. (2002). Development of a uniquely South African leadership questionnaire. *South African Journal of Psychology, 32*(2), 9-25.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201-292.

Sridhar, B. S., Gudmundson, D., & Feinauer, D. (2004). Cultural assessment: Differences in perceptions between boards of directors and other organizational members. *S.A.M. Advanced Management Journal, 69*(4), 31-39.

Sterberg, R. J. (2006). Creating of vision of creativity: The first 25 years. *Psychology of Aesthetics, Creativity, and the Arts, S*(1), 2-12.

Sternberg, R. J., & Horvath, J. A. (1999). *Tacit knowledge in professional practice: Researcher and practitioner perspectives*. Mahwah, NJ: L. Erlbaum.

Stogdill, R. M. (1974). *Handbook of leadership; A survey of theory and research*. New York: Free Press.

Stone, R. (1993). A SPSS/PC+ program listing for scoring the Leader Behavior Analysis II-Self scale. *Educational and Psychological Measurement, 53*(1), 111-117.

Stricker, L. J., & Rock, D. A. (1998). Assessing leadership potential with a biographical measure of personality traits. *International Journal of Selection and Assessment, 6*(3), 164-184.

Sutton, C. D., & Woodman, R. W. (1989). Pygmalion goes to work: The effects of supervisor expectations in a retail setting. *Journal of Applied Psychology, 74*, 943-950.

Tanoff, G. F., & Barlow, C. B. (2002). Leadership and followership: Same animal, different spots? *Consulting Psychology Journal: Practice and Research, 54*(3), 157-167.

Taylor, P. J., Pajo, K., Cheung, G. W., & Stringfield, P. (2004). Dimentionality and validity of a structured telephone reference check procedure. *Personnel Psychology, 57*(3), 745-764.

Tejeda, M. J., Scandura, T. A., & Pillai, R. (2001). The MLQ revisited: Psychometric properties and recommendations. *Leadership Quarterly, 12*(1), 31-52.

Tourangeau, A. E., & McGilton, K. (2004). Measuring leadership practices of nurses using the Leadership Practice Inventory. *Nursing Research, 53*(3), 182-189.

Tracey, J. B., & Hinkin, T. R. (1998). Transformational leadership or effective managerial practices? *Group and Organization Management, 23*(3), 220-236.

Tracy, L. (1987). Consideration and initiating structure: Are they basic dimensions of leader behavior? *Social Behavior and Personality, 15*(1), 21-33.

Tziner, A., Murphy, K. R., & Cleveland, J. N. (2005). Contextual and rater factors affecting rating behavior. *Group & Organization Management, 30*(1), 89-98.

Vandenberghe, C., Stordeur, S., & D'hoore, W. (2002). Transactional and transformational leadership in nursing: Structural validity and substantive relationships. *European Journal of Psychological Assessment, 18*(1), 16-29.

Vecchio, R. P. (1990). Theoretical and empirical examination of cognitive resource theory. *Journal of Applied Psychology, 75*(2), 141-147.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*(5), 557-574.

Watson, G., & Glaser, E. M. (1942). *Watson-Glaser tests of critical thinking*. Oxford, England: World Book Co.

Weissenberg, P., & Gruenfeld, L. W. (1966). Relationships among leadership dimensions and cognitive style. *Journal of Applied Psychology, 50*(5), 392-395.

Wielkiewicz, R. M. (2002). Validity of the Leadership Attitudes and Beliefs Scale: Relationships with personality, communal orientation, and social desirability. *Journal of College Student Development, 43*(1), 108-118.

Williams, M. L., Podsakoff, P. M., Todor, W. D., & Huber, V. L. (1988). A preliminary analysis of the construct validity of Kerr & Jermier's "Substitutes for Leadership" Scales. *Journal of Occupational Psychology, 61*(4), 307-333.

Williamson, L. G., Campion, J. E., Malos, S. B., Roehling, M. V., & Campion, M. A. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology*, *82*(6), 900-912.

Witt, L. A., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology, 87*(1), 164-169.

Wodarski, J. S. (1978). The prediction of anti-social behavior: An application of regression analysis. *Corrective and Social Psychiatry and Journal of Behavior Technology - Methods and Therapy, 24*(2), 102-110.

Yukl, G. A. (2002). *Leadership in organizations* (5th ed.). Upper Saddle River, NJ: Prentice Hall.