

6-28-2010

Developing an Effective Instrument for Assessing the Performance of Public University Presidents

Dennis L. Lester

Follow this and additional works at: https://digitalrepository.unm.edu/oils_etds

Recommended Citation

Lester, Dennis L. "Developing an Effective Instrument for Assessing the Performance of Public University Presidents." (2010).
https://digitalrepository.unm.edu/oils_etds/18

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Organization, Information and Learning Sciences ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Dennis L. Lester

Candidate

Educational Leadership and Organizational Learning

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Patricia Borrie

, Chairperson

William J. Brander

[Signature]

[Signature]

Bsmall

**DEVELOPING AN EFFECTIVE INSTRUMENT FOR ASSESSING THE
PERFORMANCE OF PUBLIC UNIVERSITY PRESIDENTS**

BY

DENNIS LESTER

B.S., Aeronautical Technology, Arizona State University, 1974
MBA, Phillips University, 1980

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Organizational Learning and Instructional Technology

The University of New Mexico
Albuquerque, New Mexico

May, 2010

DEDICATION

This dissertation is dedicated to the distinctive leaders and followers I have encountered in the military, private business sector, and public education system. These individuals distinguished themselves by their ability to see more and believe more, to expect more and respect more, and to care more and share more. Most notable was their unselfish service that enabled others to expand their horizons and to achieve more than they originally thought possible.

ACKNOWLEDGMENTS

I am grateful for the support I have received over the past five years as a student at the University of New Mexico (UNM). First, I would like to thank Dr. Patricia Boverie who served as my professor for several classes, plan of studies advisor, and dissertation chair. She is the epitome of an educator – a subject matter expert, coach, and mentor who always found the time to help. It was also an honor and joy to work with the other members of my dissertation committee – Dr. William Bramble, Dr. Scott Marley, Dr. Bruce Noll, and Dr. Mark Salisbury. Whenever possible, I registered for courses taught by these professors who eventually became members of my dissertation committee because of the challenging and stimulating environments they created in and out of the classroom.

I would also like to express my appreciation to the administrators, faculty, and staff at UNM and New Mexico State University (NMSU) who opened the doors to their campuses for my research. The UNM and NMSU faculty took time out of their busy days to provide their honest opinions and insights on characteristics of successful presidents and on how university constituents can work together to improve themselves and their institutions. In particular, I am indebted to the faculty leadership at UNM and NMSU for encouraging their colleagues to participate in my study.

Finally, I would like to thank my family for their advice, encouragement, and patience along the way.

**DEVELOPING AN EFFECTIVE INSTRUMENT FOR ASSESSING THE
PERFORMANCE OF PUBLIC UNIVERSITY PRESIDENTS**

BY

DENNIS LESTER

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Organizational Learning and Instructional Technology

The University of New Mexico
Albuquerque, New Mexico

May, 2010

**DEVELOPING AN EFFECTIVE INSTRUMENT FOR ASSESSING THE
PERFORMANCE OF PUBLIC UNIVERSITY PRESIDENTS**

BY

DENNIS LESTER

B.S., Aeronautical Technology, Arizona State University, 1974
MBA, Phillips University, 1980
Ph.D., Organizational Learning and Instructional Technology, The University of
New Mexico, 2010

ABSTRACT

Conducting a worthwhile assessment of the performance of senior leaders such as university presidents poses unique challenges for public institutions of higher education. One of the most difficult issues is determining the *content* and *format* of the assessment instrument. Due to the breadth and complexity of the job, the list of potential criteria that could serve as content for the assessment instrument is almost limitless. Turning to the format, there are also many options with respect to the arrangement of the assessment instrument or plan for its administration. Based on a review of literature for this study, there does not appear to be a comprehensive approach for developing the content and format of a president assessment instrument that provides sufficient reliability and validity evidence for the ratings derived from such an instrument.

In order to fill an apparent gap in research on university president assessment, this study proposes a model for determining the content and format of an assessment instrument – the Assessment Instrument Development Approach (AIDA). The AIDA

model incorporates a mixed-methods research design aimed at identifying the content and format of an assessment instrument that will produce reliable and valid results.

Addressing the question, “*What approach can UNM and other public universities use to develop an effective performance assessment instrument for their presidents?*” this study critiques the initial AIDA model and suggests a revised model based on study findings.

Addressing the question, “*What is the preferred content and format for a president performance assessment instrument?*” this study proposed an initial framework for the relationships among variables derived from pertinent theory and concepts pertaining to university president assessment. Considering the study results, a revised framework was developed along with example hypotheses that should be tested in future research to gain additional insights into performance assessment for university presidents.

Since two earlier studies on president effectiveness indicators obtained feedback from presidents and boards of trustees, the faculty of two research universities in the southwest were chosen as the target population for this study in order to obtain the perspectives of the third body that participates in shared governance. Qualitative data was collected through 15 individual interviews, 2 focus group interviews, a pilot survey, and a final survey. Quantitative data was collected through the pilot survey and final survey. StudentVoice™ administered the web-based pilot survey for the University of New Mexico (UNM) faculty and final survey for both the UNM and New Mexico State University (NMSU) faculty. There were 106 faculty members who volunteered to complete the pilot survey and 280 faculty members who completed the final survey.

The AIDA model was an effective tool for identifying the content and format of a president assessment instrument based on the perspective of faculty members who

participated in the study. Incorporating a mixed-methods design, the AIDA model enabled the researcher to analyze the data from different perspectives and to identify complementary and conflicting findings. While the application of the AIDA model was time consuming because it included qualitative and quantitative data collection and analysis, it proved to be useful in integrating and condensing a large amount of data and in making the results understandable.

Addressing the question of the preferred content of a president assessment instrument, over 200 potential assessment criteria were identified relating to traits and behaviors of a president and performance outcomes at the university level. These candidate criteria were prioritized based on the outcome of surveys and interviews of university faculty. Exploratory factor analysis was used to identify overarching constructs to which these criteria related and to provide insight into a methodical approach to reduce the number of items in an assessment instrument to those that are the most relevant. The constructs identified in this study that relate to university president performance were *strategic leadership, consideration, continuous improvement, university mission support, interpersonal competence, stewardship, academic quality, and responsibility*.

Addressing the question of format, participating faculty members believed that a president assessment instrument should be a formalized tool that is administered according to written policies and procedures. Study results showed that faculty members preferred annual 360-degree assessments involving multiple constituents and stakeholders that focus on president development and improvement. Faculty also

identified external factors that should be considered in performing an assessment that includes qualitative and quantitative assessment criteria.

The results of this study reveal there are many candidate criteria and formats for assessing performance. What appears to be lacking in literature and in practice is a means to identify the best criteria and formats that will produce reliable, valid, and useful results for the assessment of university presidents. The methods and findings described in this study provide additional insight into the “means” for developing an assessment instrument and the “ends” which are fair, equitable, and productive assessments of university president performance.

TABLE OF CONTENTS

LIST OF FIGURES	xvii
LIST OF TABLES	xviii
CHAPTER ONE INTRODUCTION	1
Introduction to the Chapter	1
Background to the Study	1
Problem Statement	5
Example of the Problem	5
Purpose and Justification of the Study	8
Theoretical and Conceptual Framework	9
Research Questions and Hypotheses	10
Research Questions	10
Hypotheses	10
Overview of Methodology	11
Research Design and Methods	11
Levels of Measurement	12
Target Population	12
Sampling Plan	13
Benefits of the Study	13
Assumptions and Limitations	14
Assumptions	14
Limitations	15
Definitions	15

Organization of Dissertation	15
CHAPTER TWO LITERATURE REVIEW	17
Introduction	18
Background.....	18
Assessment Definitions	20
History	21
Theories and Concepts	24
Learning.....	26
Leadership.....	27
Management	33
Followership	35
Organization and Performance Outcomes	36
Performance Assessment.....	39
External Factors.....	43
Attribution.....	45
Demographics	45
Economics.....	46
Followers	46
Organization	46
Politics	47
Raters	47
Assessment Instrument Development	49
Content and Format of Assessment Instruments.....	51

University Considerations	54
Background.....	55
Assessment Instrument	58
Methodology, Research Design, and Methods.....	71
Evaluation of Previous Research.....	86
Supporting Evidence.....	86
Contradictory Evidence	89
Gaps in Knowledge.....	91
Justification for Current Study	92
CHAPTER THREE METHODOLOGY	94
Theoretical and Conceptual Framework	94
Problem Statement.....	95
Research Questions.....	95
Hypotheses.....	95
Methodology, Research Design, and Methods.....	97
Rationale for Methodology, Research Design, and Methods.....	97
Assessment Instrument Development Approach	98
Data Management and Analysis.....	103
Data Files and Data Coding	104
Data Collection.....	105
Instruments	105
Sources.....	106
Purposes.....	106

Procedural Details.....	108
Data Verification.....	113
Data Validation	114
Trustworthiness of Qualitative Data.....	115
Reliability and Validity of Quantitative Data.....	116
Validity of Data in Mixed-Methods Research.....	121
Data Analysis	122
Qualitative Data Analysis.....	122
Quantitative Data Analysis.....	127
Mixed-Method Analysis.....	133
Risks and Mitigations.....	134
Summary	135
CHAPTER FOUR RESULTS	137
Problem Statement	137
Research Questions	137
Restatement of Assessment Instrument Development Approach	138
Restatement of Preliminary Theoretical and Conceptual Framework	139
Findings on Assessment Instrument Development Approach	141
Findings from Preliminary Analysis.....	142
Archival Data and Literature Review	142
Critical Incident Technique and Grounded Theory.....	147
Findings from Intermediate Analysis	147
Interviews	148

Pilot Survey	149
Data Analysis.....	151
Findings from Final Analysis	152
Integration of Qualitative and Quantitative Data Analyses.....	153
Assessment of Reliability and Validity	153
Presentation of Study Results	154
Summary of Findings on Assessment Instrument Development Approach	155
Findings on Assessment Instrument Content and Format.....	156
Findings from Preliminary Analysis.....	156
Critical Incident Technique and Grounded Theory	158
Findings from Intermediate Analysis	160
Pilot Survey Findings	167
Integration of Qualitative and Quantitative Findings from Intermediate Analysis.....	175
Summary of Findings from Preliminary and Intermediate Analysis.....	180
Findings from Final Analysis	183
Final Survey Quantitative Findings.....	184
Final Survey Qualitative Findings.....	197
Final Interview Findings.....	198
Integration of Qualitative and Quantitative Findings from Final Analysis .	199
Summary of Findings on Assessment Instrument Content and Format	206
Findings from Reliability and Validity Assessment	208
Qualitative Dependability and Validity Findings	208

Quantitative Reliability and Validity Findings	213
Summary of Findings from Mixed-Methods Reliability and Validity Assessment.....	218
CHAPTER FIVE DISCUSSION.....	219
Assessment Instrument Development Conclusions	219
Conclusions on Assessment Instrument Development.....	220
Convergent Conclusions on Assessment Instrument Development	223
Assessment Instrument Content and Format Conclusions	231
Implications of the Study	239
Theoretical Implications	239
Research Implications.....	240
Applied Implications	241
Limitations of the Study.....	242
Methodology.....	242
Reliability and Validity.....	243
Recommendations	244
Summary of Discussion	248
APPENDICES.....	251
APPENDIX A DEFINITION OF TERMS.....	252
APPENDIX B INDIVIDUAL INTERVIEW MATERIALS.....	262
APPENDIX C FOCUS GROUP INTERVIEW MATERIALS.....	268
APPENDIX D PILOT SURVEY	274
APPENDIX E FINAL SURVEY	281

APPENDIX F CONSENT FORMS.....	289
APPENDIX G DATA COLLECTION FORMS.....	294
APPENDIX H INSTITUTIONAL REVIEW BOARD (IRB) APPROVALS.....	297
APPENDIX I INITIAL INTERVIEW FINDINGS	306
APPENDIX J PILOT SURVEY FINDINGS	321
APPENDIX K FINAL SURVEY FINDINGS	325
APPENDIX L FACTOR ANALYSIS FINDINGS	332
REFERENCES.....	336

LIST OF FIGURES

Figure 1. Preliminary Theoretical and Conceptual Framework.....	10
Figure 2. Structure of Literature Review	17
Figure 3. Preliminary Theoretical and Conceptual Framework.....	96
Figure 4. Assessment Instrument Development Approach.....	99
Figure 5. Data Management and Analysis Process.....	104
Figure 6. Grounded Theory Procedures.....	125
Figure 7. Statistical Testing Process	128
Figure 8. Example Linkage between Qualitative and Quantitative Analysis	134
Figure 9. Assessment Instrument Development Approach.....	138
Figure 10. Preliminary Theoretical and Conceptual Framework.....	140
Figure 11. Proposed AIDA Model.....	245
Figure 12. Proposed Theoretical and Conceptual Framework.....	246

LIST OF TABLES

Table 1. Stakeholders.....	69
Table 2. Data Collection	107
Table 3. Risks and Mitigations	135
Table 4. Highest and Lowest Scores for Traits, Behaviors, and Performance Outcomes	170
Table 5. Highest to Lowest Scores for External Factors	172
Table 6. Highest to Lowest Scores for Assessment Instrument Format	173
Table 7. Ratings of Pilot Survey Items by Category	173
Table 8. Integrated Findings on Traits, Behaviors, and Performance Outcomes	177
Table 9. Integrated Findings on External Factors that Affect Performance Ratings	178
Table 10. Integrated Findings on Performance Assessment Instrument Format	179
Table 11. Demographic Profile of Survey Participants	185
Table 12. Factor Scales	192
Table 13. Factor Correlations and Factor Alpha Coefficients	193
Table 14. Top 25 Scores for Candidate Assessment Instrument Content.....	194
Table 15. External Factors that Could Impact President Performance Ratings.....	196
Table 16. Assessment Format	196
Table 17. Top 25 President Assessment Traits and Behaviors Criteria.....	200
Table 18. Top 25 President Assessment Performance Outcome Criteria.....	202
Table 19. Top 25 External Factors Influencing Performance Assessment Ratings.....	203
Table 20. Top 25 Characteristics of an Effective President Assessment Format	204
Table 21. Top 25 Assessment Instrument Content Items	235

Table A1. Comparison of Terms for this Study.....	261
Table E1. Final Survey Coding.....	287
Table I-1. Traits of a good university president.....	308
Table I-2. Positive behaviors of a university president.....	310
Table I-3. Negative behaviors of a university president	312
Table I-4. Preferred performance outcome measures	314
Table I-5. Factors beyond president’s control that can affect performance ratings.....	316
Table I-6. Value of formal assessment to the president and university	317
Table I-7. Purpose of university president assessment	319
Table I-8. Recommended participants in university president assessment.....	320
Table J-1. Pilot survey descriptive statistics	322
Table K-1. Recommendations from Final Survey on Candidate Assessment Criteria..	326
Table K-2. Recommendations from Final Survey Participants on External Factors	328
Table K-3. Recommendations from Final Survey Participants on Assessment Format	328
Table K-4. Final Survey Comment Codes and Counts from Atlas TI®	329
Table L-1. Factor Loading for Rotated Factor Structure from Responses to the Final Survey (N =280)	333
Table L-2. Scree Plot from Varimax Rotation of Final Survey Data	335

CHAPTER ONE

INTRODUCTION

Introduction to the Chapter

Performing a fair and meaningful assessment of an individual's performance poses a challenge to organizations because there are many factors that can affect the reliability, validity, and utility of the ratings, or scores, from the assessment. Conducting a high-quality assessment of chief executive officers (CEOs), such as the president of a university, has unique challenges because it is difficult to identify the criteria to include in the assessment instrument for such a complex job. This dissertation presents a mixed-methods study of university president performance assessment. The focus of this study is on the approach to developing the content and format of an assessment instrument that will provide reliable and valid results. The first chapter of this paper presents the following: 1) background to the study, 2) problem statement and problem characterization, 3) purpose and justification of the study, 4) theoretical and conceptual framework, 5) research questions and hypotheses, 6) overview of methodology, 7) assumptions and limitations, and 8) organization of the dissertation.

Background to the Study

Authors in the field of higher education point out that approaches to assessing success in universities are underdeveloped, unclear, and imprecise (Alfred, 2006) and that poorly-conceived president assessment approaches can lead to negative consequences (Michael, Schwartz, & Balraj, 2001). Davis and Davis (1999) contend that president assessments can make a difference if they are well defined, "fair in terms of expectations from often divergent constituent groups, and focus on maximizing the ability of the

president to improve the institution” (p. 119). Other education experts assert that successful models for governance of universities must place attitudes, values, and expectations of stakeholders at center stage (Gayle, Tewarie, & White, 2003). These opinions suggest that universities can improve future performance by having a systematic president assessment system in which appropriate stakeholders are involved in the process.

Although there is a wealth of anecdotal information and expert opinion on the characteristics of successful university presidents, as well as criteria and instruments various universities use to evaluate presidents, there appears to be limited research on the subject. In 1985-1986, Fisher and Tack (Fisher, Tack, & Wheeler, 1988) administered a survey of 485 administrators of 28 private foundations, 35 scholars of higher education, and more than 400 randomly selected presidents of two- and four-year, public and private institutions. The researchers in this study conducted quantitative analysis and found that more effective presidents “were found to: be less collegial and more distant; be more inclined to rely on respect than affiliation; be more inclined to take risks; be more committed to an ideal or vision rather than the institution; believe more strongly in the concept of merit pay; be more thoughtful than spontaneous; work longer hours, and be more supportive of organizational flexibility than rigidity” (p. viii). In the 2000 time frame, Kent State University conducted a study involving over 600 trustees from higher education institutions in Ohio to obtain opinions on indicators of president effectiveness (Michael, Schwartz, & Balraj, 2001). The researchers in this study provided quantitative evidence that maintaining good relationships with stakeholders were the best indicators

of president success, which appears to be somewhat contradictory of the Fisher and Tack survey results.

The Fisher and Tack pilot survey had a relatively large sample size. Their pilot survey included 109 items that were reduced to 40 items in the final survey after performing frequency and factor analysis. The target population for their survey was relatively broad as it included administrators, presidents, and scholars. While the Kent State University survey involved a relatively large number of trustees in Ohio, it appears to have addressed a relatively small set of assessment criteria (i.e., 16 items in the survey) and only one stakeholder group (members of the board of trustees) (Michael, Schwartz, & Balraj, 2001). A preliminary investigation of president assessment instruments from representative institutions revealed significant differences in the numbers and types of assessment criteria compared to the survey conducted by Fisher and Tack and the survey conducted by Michael, Schwartz, and Balraj. Additionally, there was no evidence of survey research involved a third group that participates in the shared governance of higher education institutions – the faculty.

Information is available on the subject of university president assessment; however, there appears to be limited, anecdotal data on indicators of president and institutional success at universities. This is consistent with Fisher, Tack and Wheeler (1988) and Kerr's (1984) observations that most information on presidents relates to the roles, functions, and relationships that should exist between them and the board of trustees. Former university presidents such as James Duderstadt (Duderstadt & Womack, 2003) and Frank Rhodes (2001) wrote informative books that include insightful perspectives on leadership qualities and university success indicators. Organizations

conduct periodic surveys on university and president performance that serve as excellent sources of information that can be used to develop assessment systems in a university. For example, *The Chronicle of Higher Education* (June, 2007) published the highlights of a recent survey titled, “The American College President.” The American Council of Education (ACE) conducted this survey during 2006. The full survey report (Center for Policy Analysis, 2007) provided a more in-depth view of American college presidential demographics, roles and responsibilities, and task priorities. Each of these sources offers information that can be used to develop performance assessment criteria, but does not actually recommend specific criteria.

Additional studies on presidential assessment have been conducted by various individuals and organizations. The most prominent studies appear to be those conducted by the Association of Governing Boards of Universities and Colleges (AGB) (Ingram & Weary, 2000; Nason, 1997; Schwartz, 1998). The publications that stemmed from these studies contain valuable information that universities and colleges can use to implement effective assessment systems for their presidents. The most comprehensive publication found on presidential assessment is the dissertation manuscript *Assessing the Performance of Academic Presidents* written by Merrill Schwartz (1998). Schwartz is currently the Director of Special Projects for AGB.

Two of the questions Schwartz posed in her dissertation were, “What does it mean to be an effective leader or exemplary president?” and, “How would one rate American college and university presidents?” (p. 265). Schwartz went on to say that much more is known “about procedures by which academic presidents are evaluated, but there is no agreed-upon yardstick by which we can measure their excellence” (p. 265). Based on

these observations and comments, Schwartz suggested that, “more complex definitions of effective leadership and measures of those qualities are needed” (p. 266). Schwartz’s implication is that an approach for developing these definitions and measures is lacking.

Problem Statement

A comprehensive approach is not available that universities can use to develop the content and format of a president assessment instrument that provides sufficient evidence of the reliability and validity of the ratings or scores derived from this instrument.

Example of the Problem

The University of New Mexico (UNM) Board of Regents, hereafter referred to as the *Board* in this chapter, is responsible for the governance of this major research university (*UNM Board of Regents’ Policy Manual*, 2004). In this capacity, the Board appoints the president who serves as the CEO and has the vested responsibility for the operation and management of the university. The Board’s policy manual outlines the responsibilities of the president and requires that the president report annually on the state of the university. While one section of the policy manual provides guidelines on the appointment of the president, there is no specific mention of presidential assessment. The *University Business Policies and Procedures Manual* (“The Red Book”) (University of New Mexico, 2002) contains detailed sections on performance review, recognition, and career development; however, these sections do not apply to the president.

Based on informal interviews in 2006 with a past UNM president and a past UNM faculty senate president, the university did not have a formal assessment process for the president at that time. According to these two sources, UNM presidents have prepared a list of goals and objectives on an annual basis and have provided a written statement to

the Board describing their achievements with regard to these goals and objectives.

According to one of these interviewees, the Office of the President was considering a new assessment instrument for senior administrators similar to the one being used for deans of the various colleges at UNM.

In August 2007, the Board issued the *Regents' Goals and Evaluation Criteria for the President of the University of New Mexico* (University of New Mexico, 2007b). This document, now referred to as the *President's Work Plan for FY10* (University of New Mexico, 2009), contains major goals agreed upon by the Board and the president. However, this document does not contain detailed individual performance criteria that are frequently used by public universities to evaluate their presidents. Instead, the Office of the President developed a document titled *Key Dashboard Indicators of Progress Toward UNM Presidential Work Plan AY 2007-2011* (University of New Mexico, 2008). This dashboard contains “stoplights” for quarterly reports of progress toward achieving each “milestone/benchmark” contained in the Regents’ goals and evaluation criteria document.

In addition, the goals and evaluation criteria document states that the university will track a “core set of high-level measures of performance excellence for UNM that will serve as the University’s ‘ledger system’ for evaluating and communicating performance on an ongoing basis” (University of New Mexico, 2007b, p. 2). Furthermore, the document states, “The ledger should also serve to align performance standards and accountabilities of individual colleges with those of the overall system” (p. 2). On examination, it appears this ledger contains many of the criteria that *U.S. News and World Report* (2008) uses for annual rankings of higher education institutions such as student selectivity (e.g., college entrance examination score average), faculty resources

(e.g., percentage of full time faculty), graduation rate, retention rate, and financial resources.

Using major goals of a university as indicators of president performance raises questions about reliability and validity. There are external factors that can affect the achievement of university goals aside from the president. For example, uncontrollable factors such as the cultural, economic, and political environment of the university may affect high-level outcomes. It is possible that these external factors could significantly moderate the effects of having a president who displays leadership and management traits and behaviors that well exceed stakeholder expectations, while the overall measures of university success fall short of stakeholder expectations.

Another issue of having assessment criteria that only includes university performance outcomes is the loss of the opportunity of using other measures that produce more reliable and valid ratings and more useful feedback to the president, his or her constituents in shared governance, and other university stakeholders. Research of performance instruments of several public universities reveals that their presidents are evaluated on a broad range of criteria that can be tied directly to the president to include knowledge, skills, and abilities related to leadership, management, marketing, and public relations.

According to a recent article in the *Albuquerque Journal* newspaper, the president of the Board of UNM confirmed that the university president's annual assessment was based on goals and milestones established last year (Salazar, 2009). The article went on to say that the president of the Board stated the latest evaluation of the president was conducted by the Regents ("we did it the way we've always done it") and that the faculty

and staff were not involved in the evaluation process (p. C2). Not having other constituents who are involved in shared governance (i.e., faculty, deans, department chairs, and the president) and other key stakeholders (e.g., staff and students) gives rise to additional questions as to the reliability and validity of the ratings since the Board is a relatively small group.

Purpose and Justification of the Study

The purpose of this research is to define an approach that can be used by UNM and other universities to develop the content and format of an effective president performance instrument. Since there is substantial evidence that the outcome of presidential assessment can be of great benefit to the president, it is important that universities have an effective approach for developing a useful assessment instrument. In support of this research, the boards of regents and offices of the president for two state university systems in the southwest United States (US) and five additional public universities in the western US were contacted to learn more about their president assessment systems. While preliminary research reveals there are differences between the state systems and individual public university systems, each appears to have assessment systems for the presidents that include written policies, procedures, assessment criteria, and/or assessment instruments. The results from this study will identify an approach for developing the content and format of an assessment instrument to conduct effective performance assessments for public university presidents. This study will apply this approach to define candidate criteria to include in an assessment instrument. The analysis and recommendations from this study will take into account relevant theories and counter-theories; the expectations of key stakeholder groups; the

strategy, goals, and objectives of universities; and the characteristics of performance assessment systems used at representative public universities.

Theoretical and Conceptual Framework

The theoretical and conceptual framework (Thorndike, 2005) for this study includes both internal and external models that define the overarching area of interest – assessment of university president performance. This framework was derived from research on assessing senior leaders in various types of organizations including universities. The internal model consists of the constructs associated with president performance assessment and their underlying dimensions. The internal model for this study includes the following constructs (example dimensions are in parentheses): 1) learning factors (traits/behaviors), 2) leadership factors (traits/behaviors), 3) management factors (traits/behaviors), 4) followership factors (traits/behaviors), 5) organization factors (traits/behaviors), 6) performance assessment factors (format), 7) external factors (attribution, culture, etc.), and 8) organization-level performance outcomes (student quality, student success, etc). The external model defines the expected relationships between the constructs of the internal model. Figure 1 shows the preliminary external model that illustrates the relationship between the various elements of the internal model that will be refined during this study. Since this study incorporated an exploratory research design, this model was refined over the course of the study. Chapter 5 of this document contains a description of the final model derived from qualitative and quantitative analysis conducted in support of this study.

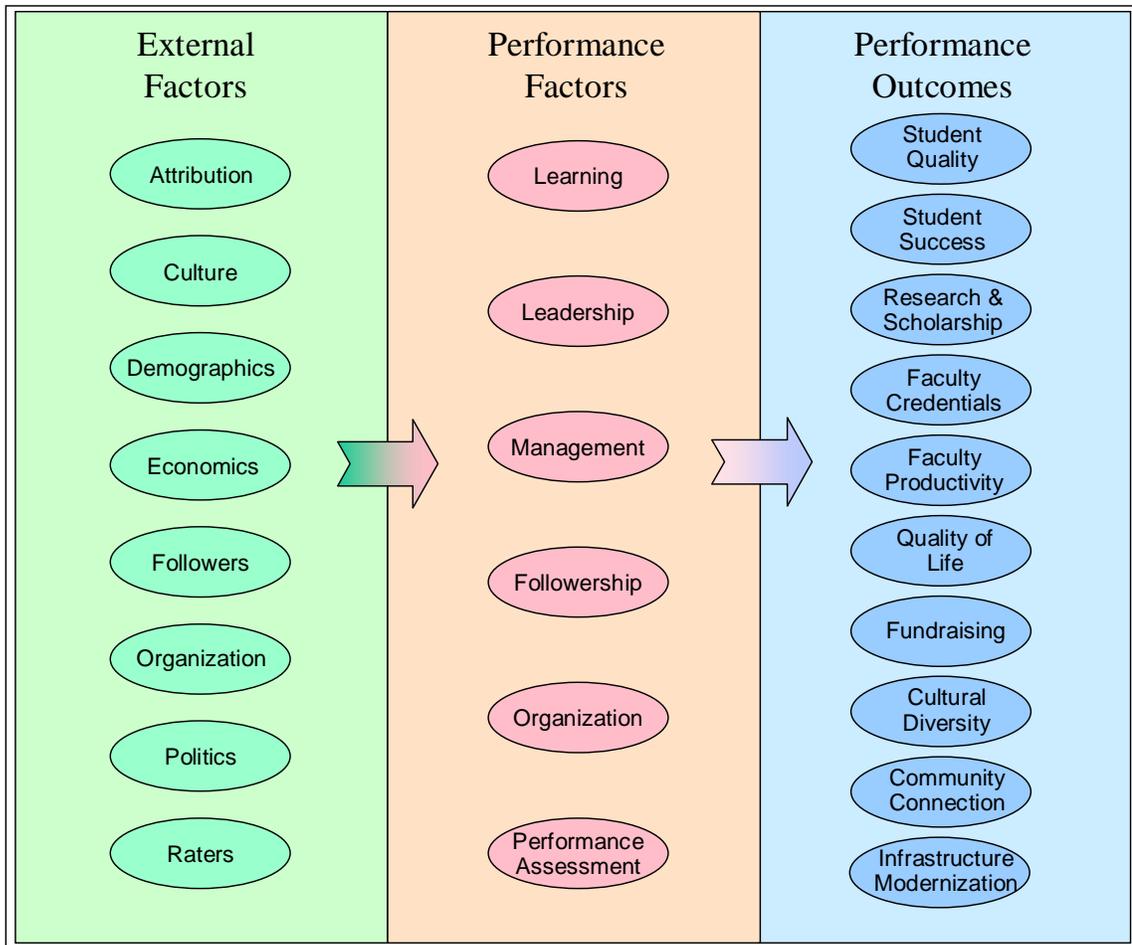


Figure 1. Preliminary Theoretical and Conceptual Framework

Research Questions and Hypotheses

Research Questions

1. What approach can UNM and other public universities use to develop an effective performance assessment instrument for their presidents?
2. What is the preferred content and format for a president performance assessment instrument?

Hypotheses

This study incorporated exploratory research methodology to identify an effective approach for developing performance assessment instruments as well as the content and

format of those instruments. Since there were many unanswered questions going into this study about the appropriate content of an assessment instrument, an exploratory research design was necessary to consolidate the large number of items (Nunnally & Bernstein, 1994) that could serve as measures for assessing the performance of university presidents. Among the main products resulting from this exploratory research are proposed hypotheses on the relative importance of various criteria that can be included in performance assessment instruments for university presidents. Chapter 5 of this document lists these hypotheses and proposes follow-up tests of these hypotheses.

Overview of Methodology

Research Design and Methods

This study consisted of a survey research design and applied mixed-methods (quantitative and qualitative) research using a combination of exploratory and triangulation techniques as described by Creswell (as cited in Fraenkel & Wallen, 2006). Quantitative data analysis involved descriptive statistics and inferential statistics. The purpose of quantitative data analysis was to identify the nature and magnitude of the relationships between variables. Qualitative data analysis involved critical incident technique (CIT) (Flanagan, 1954) and grounded theory (Glaser & Strauss, as cited in Charmaz, 2006). CIT was used for the focus group and individual interviews to identify the positive and negative traits, behaviors, and performance outcomes of university presidents. A streamlined grounded theory approach was used to identify the conceptual and theoretical constructs and dimensions that provided the basis for the items included in the survey, which provided data for the quantitative analysis.

Levels of Measurement

To analyze the relationships, the variables associated with presidential performance factors (i.e., learning, leadership, management, organization, followership, performance assessment, and performance outcomes) were treated as dependent variables (DVs). Demographic factors including age, gender, race/ethnicity, faculty member status, teaching experience, and university were treated as independent variables (IVs). The DVs were treated as interval (continuous) variables and the IVs were treated as nominal (categorical) variables.

Target Population

The target population for this study was the faculty at UNM and New Mexico State University (NMSU). The reason faculty were chosen for this research is multifold. First, according to the American College President Survey conducted in 2006 by ACE (Center for Policy Analysis, 2007), the 2,148 college and university presidents who participated in the survey cumulatively ranked faculty as their constituency that presented them the greatest challenge (p. 37). Second, this survey also revealed that the participating presidents cumulatively ranked faculty issues as one of the top ten items that occupy the most significant amount of their time (p. 38). The third reason for choosing faculty is their relatively long tenure that has given many members the opportunity to serve under multiple presidents so they have a basis for comparison. The fourth reason is that previous survey research available on president effectiveness characteristics involved only members of boards of trustees/regents, presidents, administrators, and a limited number of higher education scholars. Finally, the American Association of University Professors (AAUP) (2006) calls upon faculty members to have a primary voice in the

periodic review of academic administrator performance. Giving the faculty an opportunity to participate in this study allows them to provide valuable input on the content and format of an assessment instrument.

Sampling Plan

This study included a cross-sectional survey (Fraenkel & Wallen, 2006) with a census of the UNM and NMSU faculty. The sample frame (Dillman, 2007) is the UNM and NMSU faculty who subscribe to the all faculty list serves at each institution. The desired sample size (n) was 420 based on having 10 subjects per variable for principal components analysis (PCA) (Nunnally, 1978). Based on the results of the pilot survey and the final survey, this desired sample size was sufficient for performing multivariate analysis of variance (MANOVA) based on the specified variables, effect size, power, and confidence level (Cohen, 1988). The survey instrument was available in web-based format and in paper format as a backup. Student Voice™ (2008) administered the web-based survey. To mitigate the risks of nonresponse, Dillman's *Tailored Design* approach (2007) was applied that involved multiple follow-up contacts with the survey subjects.

Benefits of the Study

Conducting this study will have several potential benefits to UNM and other universities. First, key constituent/stakeholder groups at UNM and NMSU were involved in the study, which increases the likelihood of their acceptance and ongoing use at the universities. Second, this research examined the assessment process through a number of philosophical lenses to include postpositivism, constructivism, and pragmatism by using multiple research designs and multiple research methods. Looking at the issue of assessment from a number of perspectives increases the probability that the products will

meet the needs of the stakeholders. Another implication of this research is that the results will be transferable to the assessment systems of other senior leaders at UNM and other universities to include the provosts, vice presidents, college deans, and department heads. Having an integrated and coherent assessment system for senior leaders could increase the chances of advancing common goals, objectives, and processes to the benefit of the university and the community. Finally, this study will add information to a knowledge base that universities can use to conduct fair and meaningful assessments that can contribute to individual and institutional success.

Assumptions and Limitations

Assumptions

One of the fundamental assumptions of this research was that access would be given to UNM and NMSU faculty members for the cross-sectional survey. Another assumption was that the UNM and NMSU Offices of Institutional Research would provide demographic data of the faculty to include percentages of members by age group, gender, race/ethnicity, status as a faculty member (e.g., full-time, part-time, assistant professor, associate professor, professor, instructional faculty, visiting faculty, clinician educator, temporary faculty, and years of experience at UNM and NMSU). A final assumption was that the response rate for the individual and focus group interviews, pilot survey, and final survey would be sufficient to perform desired qualitative and quantitative analyses as part of a mixed-methods research approach. Each of these assumptions was met for this study.

Limitations

This research did not capture data from other individuals and groups that share in the governance of the university (i.e., the president and members of the board of regents) and stakeholders who have an interest in the university such as the 26,000-plus students at UNM main and branch campuses (University of New Mexico, 2010) and 34,000-plus students at the NMSU main and branch campuses (New Mexico State University (2009)). This factor limits the generalizability of findings to faculty members. Identifying assessment criteria that other constituents, stakeholders, and universities perceive to be important would require an additional investment. A series of follow-up, longitudinal studies could mitigate reliability and validity risks and could provide useful data to refine the assessment instrument for future consideration by UNM and other universities. Another limitation is that the cross-sectional survey was self-administered, so there is no way to guarantee that the actual participant is the desired participant. Despite these limitations, this study provides valuable information on an approach that can be applied to develop a useful performance assessment instrument for university presidents along with suggested content and format of that instrument.

Definitions

See Appendix A for definitions of terms that are relevant to this dissertation.

Organization of Dissertation

This dissertation contains four additional chapters in addition to this introductory chapter:

Chapter Two: Literature Review. This chapter provides highlights of literature that are pertinent to performance assessment of university presidents. This chapter

provides a foundation for the theoretical, conceptual, and methodological approaches for this study including optional research designs, methodology, and methods. It also describes candidate criteria and optional formats for performance assessment instruments.

Chapter Three: Methodology. This chapter elaborates on the methodology and design for the mixed-methods design used for this research. It includes qualitative and quantitative research methods, procedures for checking the tenability of the assumptions for statistical tests, and techniques establishing the reliability and validity of data collected during this research.

Chapter Four: Results. This chapter provides the findings derived from qualitative and quantitative data analysis based on the research question and hypotheses. The results of individual interviews, focus group interviews, the pilot survey, and final survey are presented in this chapter.

Chapter Five: Discussion, Conclusions, Limitations, and Recommendations. This chapter includes a discussion of the implications of the findings from this research and of the conclusions in terms of how this research relates to the previous body of knowledge on university president performance assessment. Any limitations that may have affected the results of this research or the application of these results in the future are addressed in this chapter. Finally, this chapter contains recommendations on the use of this study and on areas that require further investigation to expand the knowledge base in university president assessment.

CHAPTER TWO

LITERATURE REVIEW

This chapter highlights the literature that is relevant to this study of individual performance assessment with a focus on developing an assessment instrument in a useful format that incorporates reliable and valid criteria for a university president. Several theories and concepts provide a basis for developing assessment instruments. The major sections in this chapter are as follows: 1) introduction, 2) theories and concepts, 3) external factors, 4) assessment instrument development, 5) content and format of assessment instruments, 6) university considerations, 7) methodology, research design, and methods, 8) evaluation of previous research, and 9) justification for current study.

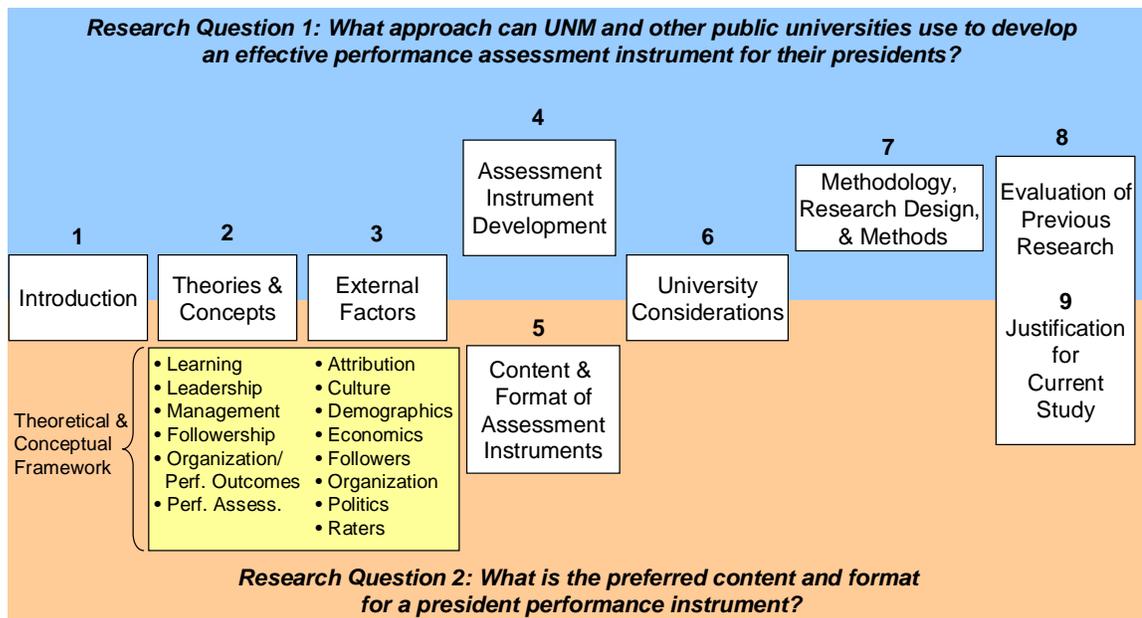


Figure 2. Structure of Literature Review

Figure 2 illustrates the structure of the sections in this literature review. Sections 2 and 3 address the preliminary theoretical and conceptual framework for the study introduced in Chapter 1 (see Figure 1). As illustrated in Figure 1, Section 4 and Section 7 exclusively

address Research Question 1 and Section 5 exclusively addresses research question 2. Each of the other sections (1, 2, 3, 6, 8, and 9) addresses aspects of both research questions.

Introduction

Background

As a component of human resource development (HRD), individual performance assessment is a widespread practice in all types of organizations. According to a survey by Locher and Teal (1988), over 90 percent of organizations conduct performance appraisals. A 2004 survey by Armstrong and Baron (as cited in Armstrong, 2009) showed that 87 percent of public and private organizations operate a formal performance management process that includes individual performance appraisals. Similarly, a survey by the College and University Personnel Association (CUPA) (as cited in Schwartz, 1998) indicated that over 80 percent of colleges and universities evaluate their presidents. Of the presidents who responded to a survey by Schwartz (1998), over 80 percent indicated that their assessments occurred on an annual basis. In the most recent American College President Study (2007) conducted by the American Council on Education (ACE) Center for Policy Analysis (2007), over 45 percent of the presidents reported that performance expectations, including the frequency of performance reviews, were agreed-upon conditions of employment.

Despite its apparent popularity and implied importance, performance assessment may be one of the least understood organizational practices and is sometimes viewed with skepticism in terms of its contribution to individual and institutional performance. Bernardin and Beatty (1984) believe that performance assessment is one of the most

neglected areas in all of HRD. Other authors suggest that designing performance assessment systems has been left to “experts in the back room” in the past (Mohrman, Resnick-West, & Lawler, 1989). Key factors that detract from the perceived or actual value of performance assessment include: 1) unclear performance criteria and ineffective rating instrument, 2) poor relationship between the rater and ratee, 3) lack of rater information on ratee performance, 4) lack of ongoing performance feedback, 5) ineffective linking of assessment to reward, 6) lack of motivation or skills on the part of the ratee, and 7) lack of focus on management development and improvement (Longnecker, 1997). Longnecker and others (Coens & Jenkins, 2000; Ingram & Weary, 2000; Nason, 1997) warn of the negative consequences associated with managerial appraisals.

Based on its widespread use and the strong endorsement it receives by many authors, individual performance assessment is a key component of HRD and can serve to improve the performance of individuals and organizations (Armstrong, 2009; Berk, 1986; Ingram & Weary, 2000; Latham & Wexley, 1994; Lecky-Thompson, 1999; Miller, R., & Miller, P., 2000; Nason, 1997; Pettijohn, R., Parker, Pettijohn, P., & Kent, 2001; Sanaghan, Goldstein, & Gavel, 2008; Schwartz, 2001). Literature also suggests that assessing performance using appropriate criteria can increase the value of performance assessment in organizations (Bernardin & Beatty, 1984; Latham & Wexley, 1994; Longnecker, 1997; Michael, Schwartz, & Balraj, 2001; Pettijohn, R., Parker, Pettijohn, P., & Kent, 2001; Nason, 1997; Schwartz, 1998; Sokol & Oresick, 1986). However, Illgen and Favero (1985) believe that researchers have been lax in mapping the social constructs related to individual performance to the assessment itself. For example, there

are interaction effects between external factors and individual behaviors that can affect performance outcomes.

Longnecker's (1997) survey of 120 seasoned managers from five large U.S. organizations found that the number one cause of ineffective managerial assessment is unclear performance criteria and an ineffective rating instrument. Nason (1997) reports that university president assessments are sometimes undertaken without prior agreement on the criteria, and this has led to "unhappiness and bitterness" on the part of presidents and trustees (p. 36). Nason also suggests that the lack of predetermined criteria that reflect the expectations from different constituencies can invalidate any criticisms of the president. This chapter reviews pertinent literature that relates to the constructs from which the assessment criteria can be derived to produce reliable, valid, and trustworthy ratings of university presidents.

Assessment Definitions

Scriven (1980) identifies several widely used terms that are synonymous with assessment including "evaluation, appraisal, critique, and review" (p. 2). Similarly, there are many definitions of performance assessment because it can be applied to different objects such as individuals, groups of individuals, programs, sets of programs, products, product lines, services, and organizations (Russ-Eft & Preskill, 2001; Scriven, 2007). The focus on this study is assessment of an individual performance – that of a university president.

Individual performance testing or assessment can be defined by job content domain, such as measuring the knowledge, skills, and abilities of an individual (American Educational Research Association, American Psychology Association, &

National Council on Measurement in Education, 2004). This study concentrates on measuring knowledge, skills, and abilities in terms of traits, behaviors, and performance outcomes associated with presidents of public universities. Additionally, individual performance assessment can be defined by its purpose or use such as individual development, compensation decisions, employment continuation decisions, and/or organizational development (Armstrong, 2009; Bruns, 1992; Davis, W. E., & Davis, D. R., 1999; Ingram & Weary, 2000; Latham & Wexley, 1994; Lecky-Thomson, 1999; Nason, 1997; Pettijohn, R., Parker, Pettijohn, P., & Kent, 2001; Schwartz, 2001). This study identifies criteria that can be applied to each of these areas to assess the performance of a university president (sometime referred to as a chancellor) who is considered the chief executive officer (CEO) of a public university.

Performance assessment can also be defined in terms of its degree of formality and frequency (Nason, 1997; Schwartz, 1998). According to Nason (1997), a formal performance assessment includes terms, conditions, expectations, systematic processes, and clearly articulated criteria for the assessment. The characteristics of a formal assessment that Nason describes are sometime referred to as policies and procedures. Nason also notes that president performance assessments are normally performed on an annual basis; however, other authors such as Ingram and Weary (2000) discuss the option of conducting more in-depth assessments on a five to ten year basis for various purposes such as confirming or adjusting longer-term institutional goals and priorities.

History

There is evidence that measurement of individual achievement has been taking place for centuries (Armstrong, 2009; Thorndike, 2005). As another form of

measurement, individual performance assessment has been conducted in America for well over 100 years (Grote, 2002). The foundation of individual performance assessment was built on psychological measurement that had its beginnings in the late 1800s (Thorndike, 2005). The industrial age that took hold in the United States in the nineteenth century brought attention to measuring worker performance to increase product output. Frederick Taylor's scientific management efforts are given credit for promoting the use of performance appraisals by companies (Grote, 2002).

During the twentieth century, pressures associated with business competition led to increased emphasis on HRD as a mechanism to improve the quality and marketability of products and services. Formal performance assessment on an annual basis became a norm for employees (Grote, 2002). Legislation during the 1960s and 1970s associated with equal employment opportunity (EEO) and federal employment influenced the formalization of performance assessment (Nathan & Cascio, 1986).

From a legal perspective, performance assessment has been used as evidence in cases involving employee discrimination and termination (Latham & Wexley, 1994) in many instances. While there may be no local and state laws that specifically address performance assessment, companies typically use formal assessment systems to support HRD and to provide a basis for employment decisions that may be contested by employees. Since performance assessment can be considered a form of a test from a legal standpoint (Nathan & Cascio, 1986), various authors recommend that organizations refer to the *Standards for Educational and Psychological Testing*, written by the American Educational Research Association (AERA), American Psychological

Association (APA), and National Council on Measurement in Education (NCME) (2004), to ensure the reliability and validity of assessment results.

While most organizations continue to use formal assessment processes and instruments (Grote, 2002), there appears to be concerns about its effectiveness in improving individual or organizational performance (Armstrong, 2009; Coens & Jenkins, 2002). Typically, an individual's supervisor is the only rater, so the feedback to the employee is limited to only one person's perspective. In addition, there are several biases that can influence the rater, resulting in errors such as halo effect, leniency, attribution, and lack of current, first-hand knowledge of the ratee's performance (Armstrong, 2009; Bernardin & Beatty, 1984; Borman, 1986; Coens & Jenkins, 2000; Grote, 2002; Latham & Wexley, 1994). To mitigate these errors, some organizations incorporate training as part of their assessment processes (Armstrong, 2009; Latham & Wexley, 1994; Mohrman, Resnick-West, & Lawler, 1989). Other organizations are incorporating self-assessment and multiple rater reviews for a more integrated approach (Armstrong, 2009; Bernardin & Beatty, 1984).

Some authors suggest that formal assessments be replaced, or at least augmented, with coaching and mentoring to avoid the negative effects of formal assessment that include dissidence, disenfranchisement, and demoralization (Armstrong, 2009; Coens & Jenkins, 2000). Coens and Jenkins recommend that assessment should be focused on groups and teams as basic operating units to make them more accountable and to mitigate the negative aspects of individual performance assessment. Many believe it is difficult, if not impossible, to tie individual performance to organizational performance because there are so many confounding factors (Armstrong, 2009; Illgen & Favero, 1985).

Armstrong's Handbook of Performance Management (2009) contains a comprehensive description of the concerns various authors have expressed about individual performance assessment. Despite the potential negative effects of individual performance assessment, an argument can be made that formal assessment, if done properly, can help individuals and organizations be more successful (Armstrong, 2009; Bernardin & Beatty, 1984; Davis, W. E, & Davis, D. R., 1999; Latham & Wexley, 1994; Mohrman, Resnick-West, & Lawler, 1989; Pettijohn, R., Parker, Pettijohn, P., & Kent, 2001; Schwartz, 2001).

Theories and Concepts

Research has shown there are several theories and concepts that relate to performance assessment including learning, leadership, management, followership, organization, and the format of performance assessment itself. With respect to learning, long-standing theories such as behaviorism, cognitivism, humanism, social theory, and constructivism have demonstrated ties to performance assessment. Recent research indicates that situational, contingency, attribution, charismatic, transformational, servant, strategic, and integral leadership theories are relevant for assessment of senior executives. This section provides a short description of these theories/concepts and examples of performance assessment criteria derived from them.

Many authors suggest that the differences between leadership and management are that leadership focuses on *what to do* and management focuses on *how to do it* (Kotter, 1998; Zaleznik, 1998). Other authors such as Drucker (2008) and Cooke and Tate (2005) believe that good managers have good leadership qualities. This study will address some of the classical functions of management including planning, organizing, directing, staffing, and controlling (Kerzner, 2006). Since the president reports to another

senior leader (i.e., the president of the board of regents or trustees) in the organizational structure and others report to the president, it is important for him or her to be a good follower as well as a good leader. Accordingly, this study examines the various dimensions of followership as defined by Chaleff (2003) and Kellerman (2008). Organization theory comes into play in performance assessment (Wexley, 1986) as well the theory associated with assessment itself. The paragraphs that follow describe various theories that may serve as a basis for determining the criteria to include in a useful assessment of a public university president.

In addition to the theories that can be used to identify appropriate performance assessment criteria, counter-theories suggest that outcomes at an organizational level (e.g., achievement of university goals and objectives) are influenced by many external factors (Armstrong, 2009; Illgen & Favero, 1985) that are potentially beyond the control of the president. For example, the number of years of experience of a faculty member (external factor) may influence his or her perception of the importance the president's knowledge of new technologies (trait) or the level of funding the university receives for research (performance outcome). The following paragraphs in this section describe the more contemporary theories and counter-theories that can be applied to developing an effective performance assessment instrument for university presidents. Using Armstrong's (2009) terms of reference, the counter-theories would be referred to as *contextual factors* or *environmental factors* that "strongly influence the content of performance management procedures, guidelines, and documentation and the processes that make it work" (p. 259). This study refers to the counter-theories as *external factors*.

Learning

In examining learning theories, behaviorism is associated with providing a certain stimulus (such as monetary reward) to motivate a desired response (such as improved performance) (Merriam & Caffarella, 1999). Incorporating behaviorism theory, the performance assessment instrument may include criteria that explore how well the president provides positive reinforcement to individuals who perform well.

The learning theory of cognitivism relates to information processing and knowledge development (Driscoll, 2000). For example, knowledge and intellect are cognitive traits that may be included as performance criteria in the assessment instrument. Having the ability to retrieve and apply knowledge to new problems is an example of a cognitive behavior (Knowles, Holton, & Swanson, 2005). Another aspect of cognitivism is applied when models or constructs are used to structure a body of knowledge (Driscoll, 2000) such as defining the categories of performance criteria that should be included in the assessment instrument. Including assessment criteria that address sharing information among executive team members relates to the theory of shared cognition (Cannon-Bowers & Salas, 2001). Shared cognition can also be leveraged when the assessment instrument is co-developed and implemented by the rater, ratee, and other appropriate stakeholders so everyone has a common understanding of the performance criteria. Situated cognition (Driscoll, 2000) can be assessed by including criteria associated with how well an individual adapts to changes. The assessment instrument can also leverage the theory of situated cognition (Driscoll, 2000) when the assessment criteria are developed in the context of external factors, such as the political and cultural aspects of the environment that can affect performance outcomes.

Humanist theory and social theory (Merriam & Caffarella, 1999) can serve as an underpinning to performance assessment in terms of motivating the ratee to achieve various levels in Maslow's hierarchy of needs and maintaining healthy relationships with others, respectively. The ability of an individual to provide a learner-centered environment in their organization is an example of assessment criteria derived from the theory of humanism. From the standpoint of social theory, the ability of the ratee to interact effectively in the social environment, internal and external to the organization, may be a key rating factor. Goleman (2006) refers to the ability to maintain healthy interpersonal relationships as *social intelligence*.

Examples of rating factors derived from constructivism theory (Merriam & Caffarella, 1999; Mezirow & Associates, 2000) are the ratee's effectiveness in creating opportunities for experiential learning and meaning-making to resolve problems, to innovate, and to foster continuous improvement within the organization. Using a bottom-up, inductive reasoning approach such as critical incident technique (CIT) (Flanagan, 1954) and grounded theory (Charmaz, 2006) to develop the constructs and hypotheses that apply to a domain of knowledge such as assessment is another practical application of constructivism.

Leadership

In general terms, Kotter (1990) defines leadership as a process associated with establishing direction, aligning people, motivating people, and inspiring people. He believes that a leader establishes direction by developing a vision and strategy for producing change to achieve that vision. Alignment of people is achieved when they understand the vision and they become committed to its achievement. Through

motivation and inspiration, the leader “keeps people moving in the right direction despite major political, bureaucratic, and resource barriers to change by appealing to basic and often untapped needs, values, and emotions” (p. 5). Kaiser, Hogan, and Craig (2008) define leadership in terms of three fundamental qualities: 1) having the ability to influence individuals’ willing contribution to the good of the group, 2) having the ability to coordinate and guide a group to achieve its goals, and 3) having the ability to focus group performance toward the achievement of organizational goals with due consideration of competition for scarce resources. The various leadership theories that follow build on Kotter’s (1988) and Kaiser, Hogan, and Craig’s (2008) definitions of leadership and provide a basis for additional criteria that can serve as a basis for assessing the performance of university presidents.

Trait, situational, contingency, attribution, charismatic, transformational, servant, strategic, and integral leadership are among the more contemporary theories that may be applicable to the assessment of university presidents. A significant amount of research has been conducted on leadership trait theories in an attempt to identify normative personal characteristics of leaders (Armstrong, 2009; Bass, 1990; Hogan, R., Curphy, & Horgan, J, 1994). Much of the evidence pertaining to trait theory was obtained through 52 factorial studies conducted between 1945 and 1970 (Bass, 1990). Since that time, several personality psychologists have endorsed a “big five model of personality structure” (Hogan, R., Curphy, & Horgan, J, 1994; Moberg, 1999) that includes reliable characteristics of leaders: 1) extroversion, 2) agreeableness, 3) conscientiousness,

4) neuroticism, and 5) openness. Each of these five personality factors with the appropriate degree of elaboration could serve as useful criteria in a performance assessment.

Bass (2008) contends that integrity is at the core of character and ethical leadership. He cites studies that showed a relationship between integrity and the personality traits of extroversion, agreeableness, and conscientiousness as well as other virtues such as authenticity, honesty, and truthfulness. Gruder (2008) reinforces Bass' perspective on the importance of integrity and suggests that integrity requires that an individual: 1) exhibits self-responsibility, 2) uses power fully, compassionately, and wisely, and 3) maintains alignment between his or her intentions and actions. Gruder proposes that "integrity is the wholeness that comes when we are fully authentic as individuals, compassionate and effective co-creators with others, and servants of the collective highest goals" (p. 41).

While the traits or behaviors associated with virtues, such as integrity, are potential criteria for assessing university presidents, they are value-laden, so it may be difficult to provide objective and quantifiable feedback on performance in these areas (Armstrong, 2009). One way to make these types of assessment criteria more objective is to operationalize them. For example, Bass (2008) defines authenticity in leaders as: 1) being true to themselves and others, 2) doing what they say they will do, 3) accepting responsibility for their personal and organizational actions, 4) accepting feedback from others, and 5) learning from mistakes. Gruder (2008) defines integrity as 1) adhering to an ethical code, 2) being unimpaired and sound, 3) being whole and complete, 4) doing unto others as you would have them do unto you, and 5) saying what you mean and doing

what you say. Including specific examples of behaviors, such as those associated with authenticity, in an assessment instrument should increase the reliability and validity of performance assessment ratings.

Kotter (1988) provides the following list of personal attributes for effective leadership in senior management jobs in complex business settings: 1) broad knowledge of the organization (key players, culture, history, and systems) and of the industry (market, competition, products, and technologies), 2) broad set of relationships in the organization and in the industry, 3) excellent reputation and strong track record in a broad set of activities, 4) keen mind (strong analytical capability, good judgment, capacity to think strategically and multidimensionally), 5) strong interpersonal skills (working relationships quickly developed, empathy, ability to sell, and sensitivity to people and to human nature), 6) high integrity (all individuals and groups are valued), and 7) motivation (high energy level and strong drive to lead, backed by self-confidence). While these attributes were defined by Kotter in a business context, many appear to be applicable to senior leaders in all types of organizations, including universities.

Situational leadership is based on the idea that a leader should tailor his or her leadership style (Bass, 1990) to environmental factors such as the nature of the problem and the social, political, and cultural environment of the organization. Demonstrating the ability to adapt leadership styles to different situations is an example of an assessment criterion associated with situational leadership. Having the ability to adjust plans effectively when faced with uncontrollable factors, such as the economy, is another example of situational leadership. Being open to novel ideas and approaches and being adaptable to multiple demands, shifting priorities, and change are examples of what

Goleman (2000) refers to as *emotional intelligence* that has applicability in the leadership domain.

Contingency theory (Fiedler, 1967) is related to situational leadership, but it is focused on either performing a task or on nurturing the relationships between individuals who must perform assigned tasks. Being able to strike a balance between attention to tasks and attention to relationships between coworkers is an example of a criterion that taps into the dimension of contingency theory. Having the ability to listen and empathize with a coworker prior to making a decision are examples of social intelligence (Goleman, 2007) that can be applied to maintain a good balance between tasks and relationships.

Attribution theory (Green & Mitchell, 1979) posits that leaders tailor their styles based on what attributes their followers believe are appropriate for their leaders to possess given the situation. Referring to previous definitions, attribution theory is related to situational and contingency theories. Applying attribution theory, viable performance assessment criteria may be the observed behavior of the leader actively seeking feedback from constituents and then applying this knowledge to improve his or her leadership style.

The word *charisma* comes from the Greek word meaning “gifted” (Khurana, 2003). Charismatic leadership (Bass, 1990) relates to the set of traits and behaviors leaders exhibit that inspire and motivate other individuals to follow them. Demonstrating the ability to motivate an executive staff to promote significant change in an organization is an example of an assessment criterion that is connected to charismatic leadership.

Transformational leadership (Bass, 1990; Northouse, 2007) incorporates aspects of charismatic leadership with inspirational and intellectual stimulation. Boverie and

Kroth (2001) emphasize the importance of an individual having true passion as the key element in transformation. In 1978, Mezirow introduced the concept of transformative learning (Mezirow & Associates, 2000) that suggests individuals and organizations can transform themselves through critical self-reflection and discourse enabling them to change outdated frames of reference for more useful decision making. These frames of reference consist of assumptions, attitudes, and beliefs. Assessment criteria such as creating an environment that encourages self-examination and stimulates thoughtful discourse are related to transformational leadership and learning.

Cranton (2006) amplifies Mezirow's thoughts on transformational learning. She provides a list of the following characteristics of transformational adult educators: 1) has accurate and complete information, 2) is free from coercion and distorted self-perception, 3) has the ability to weigh evidence and assess arguments objectively, 4) is open to alternative perspectives, 5) is capable of reflecting critically on presuppositions and consequences, 6) allows equal opportunity for individuals to participate, and 7) is capable of accepting informed and objective consensus. Cranton's list of criteria may also serve as a basis for presidential assessment.

Another form of leadership that is gaining popularity is referred to as servant leadership (Greenleaf, 2002). Servant leadership shifts the focus from the followers serving the leader, to the leader serving the followers through coaching, mentoring, and clearing obstacles that stand in the way of their followers (Blanchard & Associates, 2007). Serving as a coach or mentor to the members of the executive staff is a candidate assessment criterion that relates to servant leadership.

There is a great deal of literature that stresses the importance of strategic thinking and planning as the foundations to effective tactical execution in organizations (Alfred, 2006). Strategic leadership (Blanchard & Associates, 2007) includes identifying and implementing a framework to guide the day-to-day operations of an organization or the execution of programs. An example of an assessment criterion for strategic leadership is providing a strategic framework that includes the vision, mission, values, goals, and objectives of the organization.

Integral leadership (AGB, 2006) emphasizes the importance of leaders adopting a total “systems view” in which all the various components of the leadership situation are considered. Bringing together a variety of constituents and stakeholders to develop, incorporate, and evaluate the strategic direction of an organization is an assessment criterion associated with integral leadership. Literature indicates that CEOs are being held more accountable for successfully executing their strategies (Blanchard & Associates, 2007; Bossidy & Charan, 2002; Kouzes & Posner, 2007; Lencioni, 2002). The ability to apply a holistic approach to strategic planning that accounts for the complexities of the institution and the perspectives of a broad range of constituents and stakeholders is an assessment criterion that can be used to assess a president’s integral leadership skill.

Management

Some authors on leadership and management appear to use the two terms synonymously, while others point out the distinctions between the two fields of study (Drucker, 2008; Kotter, 1990; Kotter, 1998; Mintzberg, 1998; Zaleznik, 1998). One of the popular arguments that distinguishes the two disciplines is “leadership is not an affair

of the head; leadership is an affair of the heart” (Kouzes & Posner, 2007). Many authors believe that leadership is more related to the affective domain of learning and involves people’s attitudes, assumptions, beliefs, values, and emotions while management is more related to the cognitive domain that involves data processing, controlling, and thinking logically. For example, Kotter (1990) contends that leadership is a process associated with movement and change and that management is associated with consistency and order. Kotter stresses that increasing change brought on by environmental factors and complexity due to size, scope, geographic dispersion, and technology requires considerable leadership and management skill. While there is strong evidence that leadership and management are at least complementary, if not overlapping and synonymous in some authors’ minds, they are considered as separate concepts for the purpose of this study.

There are several relevant assessment criteria that can be included in an assessment construct and are associated with the traditional management functions of planning, organizing, directing, staffing, and controlling (Kerzner, 2006). Examples of assessment criteria that tap into the various dimensions of management are as follows: 1) planning (develops comprehensive execution plans to implement strategy), 2) organizing (logically arranges work assignments among staff members), 3) directing (provides a clear sense of direction for the execution of tasks), 4) staffing (recruits and retains high caliber employees), and 5) controlling (monitors appropriate performance measures to ensure accomplishment of organizational goals and objectives). Kotter (1990) reinforces Kerzner’s perspectives on management but adds budgeting (develops

plans for achieving fiscal targets) and problem-solving (develops approaches to resolve deviations from intended results) to the list of primary management functions.

Followership

One of the emerging theories related to leadership is followership (Chaleff, 2003; Kellerman, 2008). Leadership and management research has focused on the traits and behaviors that should contribute to or detract from the performance of individuals who serve in the role of followers. From Chaleff and Kellerman's perspectives, leadership research has not placed emphasis on the characteristics of good followers and how important they are to the success of their leaders, managers, and organizations. Some argue that it is impossible to be a good leader without first being a good follower (Kellerman, 2008). Furthermore, Kellerman states that everybody works for somebody – even the CEO reports to the owner of the company or the board and he or she must have good followership skills in order to be successful. While the research on followership is very limited compared to research on leadership and management, the literature suggest this emerging theory is pertinent to performance assessment.

There are several assessment criteria that are relevant to assessing the performance of an individual as a follower. Chaleff (2003) identifies five dimensions of followership in terms of having the courage to: 1) assume responsibility for themselves and the organization, 2) serve the leader and the organization, 3) challenge behaviors and policies with their sense of what is right, 4) participate in transformation by being full participants in the change process, and 5) take moral action when it is time to take a stand that is different than that of the leader's. Other examples of potential assessment criteria that relate to followership are: 1) displaying a willingness to reach consensus in situations

when there are conflicting viewpoints on priorities or appropriate actions to take in various situations, 2) demonstrating professional courtesy when dealing with constituents, and 3) promoting the goals and values of the institution. It is interesting to note that these followership factors are very similar, if not identical, to leadership factors identified previously in this literature review.

Organization and Performance Outcomes

Learning, leadership, management, and followership provide a robust theoretical foundation for assessing knowledge, skills, and abilities in the form of traits and behaviors of a university president. However, these theories do not address the desired performance outcomes at an organizational level that relate to organizational effectiveness. Various organization theories or models can fill this void by providing a basis for additional criteria to include in an assessment instrument.

Hall (1991) summarizes five organization models that can be used to develop assessment criteria. The population-ecology model is based on the idea that similar types of organizations adapt their environment through experiences and lessons learned. An assessment criterion derived from this model is having a knowledge management system that is sensitized to processing pertinent knowledge from the environment for timely decision making. The resource-dependence model is founded on the concept that no organization can generate all the resources it needs, so it proactively manipulates its environment to obtain the necessary resources. Showing the ability to choose the appropriate actions to obtain resources is a general criterion that could potentially be used in performance assessment. The rational-contingency model is based on the concept that organizations go through a process of setting logical goals and objectives to deal with the

situations they face. Establishing realistic organizational goals and objectives and successfully attaining these goals and objectives are examples of criteria that relate to the rational-contingency model.

The transactional-cost model incorporates the idea of cost-effective exchange of goods and services. This model appears to be related to the management functions of planning, budgeting, and controlling previously described in this literature review. Effective auditing and control systems incorporate Earned Value Management (EVM) processes to integrate cost, schedule, performance, and risk (Kerzner, 2006). Effective resource management is another general criterion related to the transactional-cost model that can be incorporated into performance assessment.

The institutional model pertains to the organizational structure that similar types of organizations adopt in order to remain productive and competitive. This structure can be influenced by a number of external forces such as government regulations and cultural expectations. Encouraging benchmarking to gain insight into the best practices of similar organizations is an example of a useful assessment criterion. Promoting the use of self-directed, multidiscipline, high performance teams (Boynton & Fischer, 2005) to tackle complex problems is another potential criterion that can be applied to the assessment of senior leadership.

In addition to these five organization models, Hall (1991) describes five performance models that various researchers have developed to identify measures of organization effectiveness. These models may also be used to develop organization assessment criteria. Yuchtman, Ephraim, and Seashore (as cited in Hall, 1991) conducted factor analyses of data collected from 75 insurance companies over an 11-year period and

identified several performance factors that were relatively stable for this period. Among these factors were: 1) business volume, 2) market penetration, 3) business mix, 4) productivity of employees, 5) production costs, and 6) maintenance costs. They referred to this model as the *system-resource model*. Etzioni (as cited in Hall, 1991) proposed a *goal model*, which is simply the degree to which an organization achieves its goals. The *participant-satisfaction* model (Hall, 1991) emphasizes the individual and group perspectives on the quality of the organization itself. The *social-function* model (Hall, 1991) focuses on measuring an organization's contribution to society. Finally, in recognition that organizational effectiveness is a multifaceted phenomenon, Hall's (1991) *contradiction model* that proposes that maximizing performance effectiveness in one area may compromise effectiveness in another area. For example, organizations may have multiple and conflicting environmental constraints, goals, internal constituents, external constituents, and operating timeframes. The contradiction model suggests that performance effectiveness is dependent on these multiple and conflicting factors.

These organization effectiveness models provide a basis for key performance indicators (KPIs) (Kerzner, 2006). Armstrong (2009) defines KPIs as results or outcomes that "are crucial to the achievement of high performance and provide the basis for setting objectives and measuring performance" (p. 233). Typical categories of KPIs are timeliness, cost, quality, and customer satisfaction (Kerzner, 2006). Armstrong (2009) goes on to say that a KPI measures something that is strategically important to the organization. This study uses the term "performance outcomes" instead of KPIs for the sake of simplicity. A general criterion associated with timeliness is delivering a product or meeting a goal or objective according to a schedule. A common criterion for cost is

staying on budget for a project or program. Another important performance criterion is how well a product meets predetermined specifications that include design characteristics and functionality. Satisfaction can also be measured in terms of timeliness, cost, and performance and can be applied to customers internal to the organization as well as those external to the organization.

While organization effectiveness models can be used to identify performance outcomes, Scriven (2007) warns that the achievement of goals is not necessarily the best measure of performance outcomes. Scriven contends there are several flaws in a goal-achievement model of evaluation including: 1) goals may be set too high or too low, 2) goals may be irrelevant to the needs of those being served, 3) goals may overshadow important side effects, and 4) goals may not be worth the expense in achieving them. Instead of using goals as a basis for assessment criteria, Scriven suggests that actual effects of a program or individual, and the cost of achieving those effects, serve as better measures of performance. Based on their research of 10 meta-studies on leadership measurements in past research, Kaiser, Hogan, and Craig (2008) conclude that while organizational effectiveness depends on more than leadership, the data clearly show that leaders have a substantial influence on it. Accordingly, they believe that, “Regardless of the measurement method, the unit of analysis for evaluating leadership effectiveness should be the performance of the group, team, or organization being led” (p. 107).

Performance Assessment

Compared to the previously mentioned theories and concepts, the literature available on performance assessment is much more limited, especially for senior executives in an organization. Most of the literature about performance assessment is

focused on the practical aspects of assessing employee performance. Turning to another potential source, there is extensive literature on program evaluation that is applicable to performance assessment of individuals. For example, assessment can be viewed as a form of learning (Armstrong, 2009; Russ-Eft & Preskill, 2001). Authors suggest this learning takes place through the application of a process that includes the following steps: 1) establishing performance criteria, 2) constructing standards to which the object (e.g., person or product) under scrutiny should perform, 3) measuring performance and comparing it to the standards, 4) synthesizing and integrating evidence, and 5) developing recommendations (Fornier, 1995; Russ-Eft & Preskill, 2001; Scriven, 1995). This assessment process parallels scientific method (Fraenkel & Wallen, 2006) which includes: 1) identifying a problem, 2) clarifying the problem, 3) collecting data that addresses the problem, 4) analyzing the data, 5) drawing conclusions, and 6) making recommendations. Lessons learned from systematic evaluation can be applied to a number of high-interest areas in an organization to include decision making, personal development, organizational development, and product improvement (Russ-Eft & Preskill, 2001).

Russ-Eft & Preskill (2001) and Scriven (1993) provide descriptions of the three types of evaluation – *developmental*, *formative*, and *summative*. The purpose of developmental evaluation is to define the requirements for a new program, system, or process. For instance, if an organization does not have a performance assessment system, a developmental evaluation could be performed to: 1) identify the requirements of the new system, 2) implement the system, 3) monitor execution, and 3) modify the system over time. A formative evaluation concentrates on assessing an ongoing program and on

identifying actions to improve the program (Scriven, 1993). As such, a formative evaluation is a more focused and abbreviated process compared to a developmental evaluation. The final type of evaluation is a summative evaluation. The purpose of the summative evaluation is to determine the value or worth of a program, system, or process. A summative evaluation usually results in a report that contains a final judgment or decision (Scriven, 1993) such as continuation or termination of a program. Scriven (1993) believes that the main reason for performing a summative evaluation is accountability.

Organizations can apply many of the processes and tools proposed in this document to conduct developmental, formative, or summative evaluations of systems and processes. Applying program evaluation terminology to the performance assessment of an individual, an assessment of an employee appears to be more closely associated with formative and summative evaluation rather than with developmental evaluation. The evaluation concepts described for organizations provide a starting point for determining characteristics of a performance assessment for individuals.

There are a number of purposes of performance assessment in addition to providing feedback, mentoring, coaching, and determining compensation changes (Armstrong, 2009; Grote, 2002). Some of these purposes are: 1) making a decision to retain or release an individual, 2) motivating superior performance, 3) setting and measuring goals, 4) counseling poor performers, 5) determining individual training and development needs, 6) determining organizational training and development needs, 7) providing legal defensibility for personnel actions, and 8) improving overall organizational performance. While this list is not exhaustive, it provides a sense that

organizations can use individual performance assessments for a number of reasons. It is important to determine the purpose early in the process because it has a significant influence on other design characteristics of the assessment system (Armstrong, 2009; Grote, 2002).

Another characteristic that relates closely to purpose is the frequency of an assessment. If the purpose of the assessment is to provide data for making a decision on a merit pay increase or bonus for an individual, one would expect the organization to conduct a summative evaluation process on an annual basis (Russ-Eft & Preskill, 2001; Sweeny & Manatt, 1986). If the purpose is to provide on-going feedback to an individual, the organization should consider using a formative evaluation process (Russ-Eft & Preskill, 2001; Sweeny & Manatt, 1986).

In addition to purpose and frequency, the degree of objectivity and formality are additional characteristics of a performance assessment (Bernardin & Beatty, 1984; Latham & Wexley, 1994, Russ-Eft & Preskill, 2001). An assessment that contains quantifiable criteria and standards tends to be more objective than one that contains qualitative statements about performance. Formal assessments are systematic in that they follow a specified set of policies and procedures and involve a standardized assessment instrument (Armstrong, 2009; Latham & Wexley, 1994). The results of formal assessments are sometimes made available to the public for the purpose of accountability (Nason, 1997). On the other end of the spectrum, the practice of coaching and mentoring is an example of an informal assessment (Armstrong, 2009; Latham & Wexley, 1994). Informal assessments are less systematic and do not necessarily follow any specific

procedures or format. In some cases, informal assessments are part of a day-to-day feedback process that involves a more private process involving the rater and the ratee.

Before making a final decision on the preferred characteristics of a performance assessment, an organization should consider additional factors. Some of the more important factors are the: 1) time and effort necessary to develop the criteria, standards, and instrument for the assessment, 2) individuals who should participate in the assessment process, 3) availability and use of results, 4) approach for monitoring and revising the assessment program, and 5) commitment and involvement of top management in the assessment program (Grote, 2002). Grote stresses that an assessment program will not succeed without the visible support of top management, their active participation in the process, and their demonstrated use of the results to make decisions and to set direction.

External Factors

Several external factors can influence a performance assessment aside from the knowledge, skills, and abilities of the president. For the purposes of this study, the external factors will be considered “exogenous variables” (Shadish, Cook, & Campbell, 2002, p. 507) in that they are not caused by other variables in the model. For studies investigating causation, an external factor could be considered to be a “moderator variable” (Shadish, Cook, & Campbell, 2002, p. 509) if it changes the size and direction of the relationship between a predictor variable such as a leadership trait and an outcome variable such as the amount of funding the university receives for research.

Shadish, Cooke, and Campbell (2002) suggest that moderator variables are almost always *confounded* (p. 451) with other variables, making it more difficult to determine

the relationship between a predictor variable (independent variable) and an outcome variable (dependent variable). An external factor *mediates* (p. 509) the effect if it neutralizes the relationship between the predictor variable and outcome variable. This study does not attempt to establish causation between variables or the effects of moderator and mediator variables based on collected data.

However, this study does examine external factors that can potentially affect or bias performance assessment ratings to set the stage for further research that could investigate moderator and mediator variable effects. The external factors/endogenous variables for this study are external factors that may be beyond the president's control and that could affect his or her performance assessment. These external factors could have a positive or negative impact on the ratings the president could receive on a performance assessment. Evaluators should consider these potential affects or biases to ensure an assessment is fair and equitable (Armstrong, 2009; Bernardin & Beatty, 1984; Coens & Jenkins, 2002; Grote, 2002; Latham & Wexley, 1994).

Follow-on research could benefit from the analysis of external factors as moderator and mediator variables by investigating the causal relationship between president performance ratings and high-level outcomes such as the amount of funding the university received for research. Trend analysis (Field, 2005; Fraenkel & Wallen, 2006) could also be useful for determining how president performance ratings compare the achievement of university goals and outcomes as a means for determining the appropriate content and format of a reliable and valid president assessment instrument. Research on the subject of executive leadership and performance assessment does not suggest that external factors can totally *negate* the influence that factors such as leadership and

management have on performance outcomes. Research does suggest that external factors can *impact* the effects of leadership and management factors on performance outcomes and these external factors should be taken into account when developing a performance assessment instrument and evaluating assessment results.

Attribution

Many authors believe that organizations are so complex and external factors are so numerous and influential, that it is impossible to attribute organization performance outcomes to characteristics or attributes of an individual (Bernardin & Beatty, 1984; Latham & Wexley, 1994). Overestimating the power that leaders have on organization performance is an example of fundamental attribution error (Coens & Jenkins, 2000).

Culture

There is evidence that people from different cultures prefer certain types of leaders (Gibson, 1995; Hall, 1991; Yeh, 1995; Northouse, 2007; Yu & Miller, 2005). For example, some cultures expect leaders to be charismatic, value-based, participative, team-oriented, and sensitive to people (Northouse, 2007). Other cultures may expect their leaders to be more autonomous, self-protective, and goal-oriented (Northouse, 2007).

Demographics

Demographic factors such as age, gender, position within the organization, and years of experience can affect the assessment criteria deemed important and the perceptions of how well the leader performs his or her job (Hall, 1991; Michael, Schwartz, & Balraj, 2001; Northouse, 2007).

Economics

Research provides evidence that the condition of global, national, state, and local economies can affect performance outcomes (Hall, 1991) such as raising funds that can be invested in organizational programs and infrastructure.

Followers

The characteristics and intentions of followers internal and external to the organization can affect the perceived and actual performance of a leader. For example, followers who trust and respect their leaders will be supportive of the leader, which can improve organizational performance (Kellerman, 2008).

Organization

As described previously, Hall's (1991) contradiction model proposes there are multiple competing or conflicting factors that may affect performance effectiveness. Examples of these organization factors are: 1) facing multiple and conflicting environmental constraints, 2) having multiple and conflicting goals, 3) having multiple and conflicting internal and external constituencies, and 4) operating in multiple and conflicting timeframes. Due to these factors, performance outcomes may be affected by constraints that are beyond the control of an organization or individual. Similarly, an organization or individual may take actions to improve performance in one area causing performance to regress in another area. The main implication of the contradiction model is that these factors need to be taken into account when conducting an assessment since they may affect individual performance ratings. Another implication is multiple and conflicting factors should be considered when developing the content and format of a performance assessment instrument for an individual.

Politics

Performance ratings can be biased by political considerations (Hall, 1991). For example, an individual's performance report can be affected by how well the ratee's viewpoint aligns with the parties in power such as the members of the company's board, elected officials, and appointed officials. Higher-ranking officials can also apply political pressure to achieve their goals and objectives at the expense an individual's performance in areas that others believe are of higher priority.

Raters

There are a number of factors associated with raters that can bias the results of a performance assessment. Some of these factors are as follows: 1) knowledge and experience of the rater, 2) personal relationship between the rater and ratee, 3) level of training provided the rater on performance assessment, 4) motivation level of the rater, and 5) perceptions the rater has on the qualities of a good leader. Additional factors that can affect the reliability and validity of ratings are: 1) leniency error, 2) severity error, 3) central tendency error, 4) halo errors, 5) recency error, 6) fundamental attribution error, and 7) self-serving bias (Borman, 1986; Coens & Jenkins, 2000; Latham & Wexley, 1994; Weitzel, 1987).

While organizations/raters should consider these rater errors, Latham & Wexley (1994) provide ways to mitigate their effects by using a high-quality assessment instrument and by involving peers, subordinates, and outside experts (e.g., consultants), as well as superiors, in the assessment process. Many authors refer to the use of multiple raters to include superiors, peers, and subordinates as a *360-degree assessment* or *360-degree feedback* (Armstrong, 2009; Coens & Jenkins, 2002; Fain, 2006; Grote, 2002).

The purpose of a 360-degree assessment is to increase reliability and validity since supervisor assessments (the most common form of assessment) contain rater biases (Latham & Wexley, 1994). Coens and Jenkins (2002) believe that while supervisors may be well intended, they use lenses and biases that distort individual performance. Coens and Jenkins contend that 360-degree assessment provides greater awareness of the perceptions of more people who can offer critical and useful information. Involving raters who have relatively close contact with the ratee and who have a good understanding of the ratee's responsibilities reduces bias and increases reliability and validity (Latham & Wexley, 1994). Having multiple types of raters contributes to reliability and validity because it increases the number of independent judgments (Latham & Wexley, 1994). Another key contributor to reducing bias and increasing reliability and validity is ensuring the assessment is anonymous (Latham & Wexley, 1994). The lack of anonymity can severely bias results because of the fear of retribution.

Several authors recommend that an assessment process include self-assessment (Armstrong, 2009; Grote, 2002; Latham & Wexley, 1994; Nason, 1997). Self-assessment allows the ratee to critically reflect on their performance and identify areas of improvement (Armstrong, 2009; Grote, 2002). It can also serve as a vehicle to promote discussion with the raters to compare perspectives (Armstrong, 2009; Grote, 2002) and resolve differences before determining the best courses of action for the future. Latham and Wexley (1994) stress that self-assessment promotes self-awareness that can lead to appropriate adjustments in behavior to achieve goals.

Assessment Instrument Development

Despite the vast amount of literature on the various theories associated with leadership and management and the technical aspects of developing assessment instruments, comprehensive approaches for determining the specific content and format of senior executive performance assessment instruments are not readily available. The most abundant source of information on assessment of university presidents is the Association of Governing Boards of Universities and Colleges (AGB) (Ingram & Weary, 2000; Nason, 1997; Schwartz, 1998, 2001). While the AGB publishes documents with example criteria and special considerations for assessing senior leaders such as university presidents, it does not publish detailed information on how to develop the content and format of a performance assessment instrument.

Some literature describes systematic processes for developing a performance assessment system; however, this literature does not provide specific steps for developing the assessment instrument itself. For example, Bernardin (1986) defines the following steps of a *Diagnostic Model of Appraisal System Development*: 1) assemble a task force on appraisal, 2) identify organizational variables that may have an impact on appraisal effectiveness, 3) determine the number and types of appraisal systems that appear to be feasible by examining survey results and job analysis data, 4) recommend a performance assessment system to the task force, 5) develop a prototype assessment system and propose a demonstration project, 6) conduct a demonstration project, 7) analyze the results of the demonstration project and propose changes to the assessment system based on results, 8) implement the performance assessment system, and 9) evaluate the effectiveness of the system.

Mohrman, Resnick-West, and Lawler (1989) provide a seven-step process for designing and implementing a performance assessment system. The steps of their process are as follows: 1) select the right people to be involved in the process such as human resources professionals, managers, and employees who have first hand experience with performance assessment systems, 2) decide on the process to guide the design (e.g., outside consultation, centralized development by a department within the organization, or a task force), 3) assess the current organizational situation to include the current system in place (how well it is going, what the problems are, organizational culture, organizational climate, influence/power structure, and legal requirements), 4) establish the system's purpose and objectives (e.g., performance improvement), 5) design the performance assessment system (which includes determining who appraises performance, what is meant by performance, how performance is appraised, and when is it performed), 6) experiment with the implementation through use of a dry run or pilot test and correct flaws before implementation, and 7) evaluate and monitor the system once implemented to test whether the system is achieving its objectives. This seven-step process is very similar to Bernardin's model of assessment system development described in the previous paragraph.

Bernardin and Beatty (1984) offer a flowchart for the development of performance measures to include in an assessment instrument. This flowchart includes the following steps: 1) identify candidate performance measures that can be used to accurately assess a person, 2) validate the performance measures by looking at professional standards, laws, *Uniform Guidelines on Employee Selection Procedures* (1978), and *Testing in Employment and Credentialing* (American Educational Research

Association, American Psychology Association, & National Council on Measurement in Education, 2004), 3) conduct a task- and/or behaviorally-based job analysis to identify relevant performance measurements, 4) develop a performance assessment format that considers rater differences and job characteristics and includes specific goals and feedback loops, 5) use the performance measures to support personnel decisions (e.g., promotions and training requirements), and 6) compare the relative importance of the performance measures for organizational effectiveness over time and adjust the measures as necessary. Bernardin and Beatty's flowchart, along with Bernardin's diagnostic model and Mohrman, Resnick-West, and Lawler's seven-step process provide useful insights into "what to do," however, they do not really address "how to do it."

Content and Format of Assessment Instruments

While organizations tend to standardize the content and format of the performance assessment instrument for different types of jobs, research reveals that there is minimal standardization between organizations for similar jobs. There are many reasons for these variations including differences in specific tasks performed by individuals in different organizations even though they have the same job titles (Fine, 1986; Latham & Wexley, 1994). Additional reasons for variations between organizations include the following (as alluded to earlier in this literature review): 1) purpose of the assessment, 2) perceived value of the performance assessment, 3) level of investment of resources, 4) desired degree of accountability of the rater, 5) desired degree of formality, transparency, and confidentiality in the performance assessment process, 6) and the approach for developing the content and format of the performance assessment instrument (Bernardin & Beatty, 1984; Jacobs, 1986).

Despite these differences, there are several commonalities in the content and format of performance assessment instruments. For example, with respect to the general type of content, a formal performance assessment instrument usually contains rating criteria associated with traits, behaviors, and/or performance outcomes (Armstrong, 2009; Bernardin & Beatty, 1984; Grote, 2002; Latham & Wexley, 1994; Sokol & Oresick, 1986). With regard to the general format, many organizations have instruments in the form of a written or fill-in-the-blank survey or report. The assessment instrument may contain a number of forced-response, closed-ended questions (Dillman, 2007) or statements that characterize individual traits, behaviors, or performance outcomes.

Sometimes the assessment instrument includes a *Likert scale* (Latham & Wexley, 1994) in which a numerical value corresponds to a description of an observed behavior. For example, the rating may be a “1” for a rater who never recognizes the superior performance of his or her subordinates or a “5” for a rater who always recognizes the superior performance of his or her subordinates. This popular scale is a *summated scale* (Bernardin & Beatty, 1984).

Another frequently used scale that is similar to the summated scale is the *behaviorally anchored rating scale* (BARS) (Latham & Wexley, 1994). BARS ratings are based on the perceived effectiveness of certain behaviors. For example, BARS will contain multiple statements that represent a range from effective behaviors to ineffective behaviors. The rater is asked to mark a numerical scale that coincides with the behavior statement that best describes the observed performance of the ratee. As an alternative or supplement to using quantitative rating scales, an assessment instrument may provide

space to answer open-ended questions (Dillman, 2007) that are qualitative or subjective in nature (Bernardin & Beatty, 1984).

The processes for determining the criteria to include in a performance assessment instrument vary from organization to organization. Authoritative sources suggest that organizations should first perform detailed job analyses to identify the key tasks and desired level of proficiency in performing these tasks (Bernardin & Beatty, 1984; Fine, 1984; Grote, 2002; Latham & Wexley, 1994). Researchers can use CIT to identify behaviors that contribute to and detract from job performance (Armstrong, 2009; Fine, 1986; Flanagan, 1954; Latham & Wexley, 1994; Sokol & Oresick, 1986). One of CIT's early uses was to identify performance criteria for U.S. Army Air Force pilots trained during World War II. When using CIT, it is important that the individuals who contribute to the development of performance criteria be familiar with the job. Organizations can then use the behaviors derived from CIT to develop an assessment instrument that incorporates a summated scale (e.g., Likert scale) or BARS.

Some organizations use off-the-shelf assessment instruments that are developed by commercial vendors (Connolly & Wilson, 2007; Learning Center, 2008; VisionMetrics, 2008). In some cases, organizations form an internal committee of specialists to develop performance assessment instruments (Sokol & Oresick, 1986). In other cases, organizations hire consultants to provide the assessment tool and to facilitate the assessment (Latham & Wexley, 1994). Assessment centers are businesses that specialize in assessing managerial effectiveness (Byham & Thornton, 1986; Latham & Wexley, 1994). These assessment centers use multiple assessment methods including job simulation exercises, multiple assessors, objective and subjective data collection, and

prediction of performance based on specified criteria (Byham & Thornton, 1986). Still other organizations use a combination of techniques to develop the content and format of their performance assessment instruments (Latham & Wexley, 1994).

University Considerations

There was no literature that specifically focused on developing assessment instruments for university presidents. Fortunately, there is literature on president assessment considerations and trends in the usage of presidential assessment. The dissertation by Schwartz (1998), *Assessing the Performance of Academic Presidents*, focused on procedures and consequences of performance reviews and identified gaps in knowledge on measuring president performance. Lawshe (1975) suggested “in contrast to academic achievement, the job performance universe, as operationally defined and about which inferences can be made, and its parameters, are often ill-defined, even with careful job analysis” (p. 564). Based on the writings of Lawshe and Schwartz, it appears that a limited amount of literature addresses specific criteria for assessing the performance of university presidents or a useful approach for developing these criteria.

Assessment of the president of a public university is similar in many respects to assessment of the CEO of a corporation (Schwartz, 1998). There is evidence that there are normative traits and behaviors that contribute to performance outcomes in both types of organizations (Grote, 2002; Ingram & Weary, 2000; Nason, 1997; Northouse, 2007; Sanaghan, Goldstein, & Gaval, 2008; Sokol & Oresick, 1986; Wickert & McFarland, 1967). However, authors highlight significant differences in the environments and situations in which the president of a public university and CEO of a corporation carry out their duties (Atwell, 2007; Duderstadt & Womack, 2003; Hearn & Alexander, 2006;

Padilla, 2005; Rhodes, 2001; Sanaghan, Goldstein, & Gaval, 2008).

This section provides general background on some of the unique aspects of public universities that should be taken into account when developing performance assessment instruments. More specifically, examples are provided showing the content and format of assessment instruments and considerations with regard to the stakeholders typically involved in performance assessment and their use of president assessment results.

Background

The approach to performance assessment varies from university to university except for those universities that are part of an overarching statewide system that have policies, procedures, and assessment instruments that apply to all institutions under their purview. Some universities have informal systems with no stated policies and procedures and no formal assessment instrument (Nason, 1997). Informal assessments may or may not be casual and tend to be more subjective (Nason, 1997).

Many universities in the Southwest have formal systems with descriptions of policies, procedures, and assessment instruments (e.g., Colorado State University, State of Arizona System, Nevada System of Higher Education, New Mexico State University, and University of Texas System). A search of the Internet reveals that many universities post their policies, procedures, and assessment instruments on their websites. In some cases, universities make the results of president performance assessments available to the public (Nason, 1997). In other cases, the results are confidential and only made available to a very restricted audience such as the board of regents and to the president.

Additionally, there are differences in the types of stakeholders who may participate in the assessment of the president. Sometimes the only individuals involved

in the assessment are the board of regents and the president and sometimes several constituent/stakeholder groups provide their inputs (Nason, 1997). Among the most frequently cited stakeholder groups that participate in president performance assessment are university administrative staff, faculty, and students (Atwell, 2007; Ingram & Weary, 2000; Nason, 1997; Schwartz, 1998; Schwartz, 2001).

Universities also vary in the frequency of president performance assessment. Most universities perform an assessment of their president on an annual basis (Schwartz, 1998). Other options include performing a more limited assessment of their president on an annual basis and a more comprehensive assessment on a three to five year basis (Atwell, 2007; Schwartz, 1998) and/or five to ten year basis (Ingram & Weary, 2000).

While there is a moderate amount of literature on president performance assessment in general, there is less literature that identifies approaches for developing relevant criteria that universities should include in president assessment. The AGB appears to be the most authoritative and prolific source of information on university presidential assessment criteria. The AGB document *President Assessment: A Guide for Periodic Review of Performance of Chief Executives* by Nason (1997) identifies several approaches that have been used to develop president assessment criteria in relatively general terms. For example, Nason (1997, pp. 35-36) states these criteria “emerge” in four different ways: 1) from institutional handbooks that contain general president responsibilities, 2) from the institutional goals the board and president agree upon at the beginning the president’s tenure, 3) from an agreement between the board and president after questions are raised about the need for evaluation and for benchmarks, and 4) from an invitation by the board for the president to conduct a self-assessment. Nason also

provides example assessment instruments that have been used by various universities. Some of these instruments are detailed questionnaires that contain several items and scales that apply to each item. Other instruments are more general to include letters to raters that request written feedback in designated areas of performance.

The dissertation by Schwartz (1998, p. 265) poses the following questions for further research:

1. What does it mean to be an effective leader or exemplary president?
2. How would one rate American college and university presidents?

Schwartz goes on to say there is “no agreed upon yardstick by which we can measure [president] excellence” (p. 265) and “more complex definitions of effective leadership and better measures of those qualities are needed” (p. 266). Even though the writings of Nason (1997) and Ingram and Weary (2000) offer some examples of president performance criteria, Schwartz’s research suggests that considerable work remains to be done in identifying appropriate constructs from which criteria can be defined for a fair and meaningful assessment of a university president.

Another important consideration in the assessment of university presidents is the lack of information on an effective approach that universities can use to develop an assessment instrument. One possible reason for this is that presidents and boards of regents have been more interested in the practical aspects of assessment rather than in the theoretical and situational factors related to performance assessment. This is evident in the more prominent writings on university president assessment by Nason (1997), Schwartz (1998), and Ingram and Weary (2000). Another reason that universities have not developed more effective performance assessment instruments may be due to their

reluctance to invest additional resources in the process when there are so many other competing priorities. As public universities respond to the call for greater accountability and transparency, literature suggests that there may be increasing emphasis on formal performance assessment (Ingram & Weary, 2000; Nason, 1997; Schwartz, 1998).

Assessment Instrument

Content

While there is a limited literature that addresses the content of president assessment instruments, some universities publish their policies, procedures, and assessment instruments that contain specific performance assessment criteria. These criteria include traits, behaviors, and performance outcomes that fall under the following categories identified in an AGB survey (Nason, 1997): 1) academic management and leadership, 2) administrative management and leadership, 3) budget and finance, 4) fundraising, 5) external relations, and 6) personal characteristics. An AGB publication by Ingram & Weary (2000) refined this list of criteria to include the following:

1) institutional agenda, 2) academic leadership, 3) general management and planning, 4) fiscal management and budgeting, 5) fundraising, 6) internal relationships, 7) external relationships, 8) decision making and problem solving, and 9) other perspectives (e.g., major achievements and shortcomings, closing words, etc.). Sanaghan, Goldstein, and Gavel (2008) recommend criteria such as: 1) leadership, 2) administration (including financial management), 3) donor and alumni relations (including fundraising), 4) quality of internal personal interactions, 5) strategic planning, and 6) relationship with the board of regents/trustees. Rhodes (2001) emphasizes the importance of effective application of information technology to increase access and the quality of instruction.

Alfred and Rosevear (2000) identify several guidelines associated with new organizational models for universities. These guidelines include: 1) focusing on strategies to improve competitiveness, 2) decentralizing authority by flattening the organization hierarchy, 3) empowering the staff to make decisions, 4) utilizing existing staff rather than seeking hired help, 5) encouraging team problem-solving, 6) emphasizing speed and flexibility in the business model, 7) promoting continuous learning for innovation and renewal, 8) stressing the sharing of information at all levels in the organization, 9) developing seamless relationships with customers and suppliers, 10) emphasizing customer service, 11) promoting visionary leadership, and 12) anticipating marketing needs and forecasting strategic changes to “get to the future first” (p. 17). Several of these principles are similar to those that apply to any type of organization and can serve as potential criteria for president performance assessment.

One key external factor that has an influence on university president performance is shared governance. Gayle, Tewarie, and White (2003) define shared governance in a university as “the structure and process of authoritative decision making across issues that are significant for external as well as internal stakeholders within a university” (p. 2). Their list of external and internal stakeholders includes 1) higher education associations, 2) funding organizations, 3) the US Department of Education, 4) related congressional committees, 5) accrediting institutions, 6) university system offices, 7) governors, 8) state departments or boards of education, 9) state legislators, 10) students, 11) alumni, 12) local community members, 13) trustees, 15) senior administrators, and 16) presidents (Gayle, Tewarie, and White, 2003, p. 2). Gayle, Tewarie, and White recommend that the governance model for universities in the future should place the attitudes, values, and

expectations of the stakeholders at center stage (p. 4). They also contend that the performance outcomes of the university are mediated by stakeholder attitudes, values, and expectations (p. 4).

According to Duderstadt and Womack (2003), the mediating effects of shared governance include: 1) lack of responsiveness, 2) reluctance to change, 3) distraction of senior leadership from strategic issues in favor of political agendas and personal interests, 4) lack of accountability, and 5) imbalance between authority and responsibility. Levin (2002) suggests there are several other external factors that are antithetical to the constructs of leadership in the university including: 1) scientific and scholarly criticism and skepticism, 2) authority not being found in personalities and roles, but rather in scholarship based on research, evidence, empirical data, and argumentation, 3) faculty demand for solid evidence before it is willing to change, and 4) institutional mindset of the importance of preserving its culture, pursuing knowledge, and maintaining autonomy.

Duderstadt and Womack (2003) stress the need for presidents to develop, articulate, and implement institutional visions and to serve as the university's leading recruiter in hiring talented people. The 2006 American College President Study conducted by the American Council on Education (Center for Policy Analysis, 2007) identified the following areas as those that occupy most of the presidents' time: 1) fundraising, 2) budget/financial management, 3) community relations, 4) strategic planning, 5) governing board relations, 6) personnel issues, 7), capital improvement projects, 8) facility issues, 9) enrollment management, 10) academic issues, 11) government relations, 12) crisis management, 13) entrepreneurial ventures, 14) media/public relations, 15) accountability of student learning, 16) student life/conduct

issues, 17) athletics, and 18) technology planning. This list of the main time consumers of presidents' time provides an indirect source of president performance criteria.

Starting in 1985, Fisher and Tack (Fisher, Tack and Wheeler, 1988) administered a survey of president effectiveness indicators that involved 485 participants including administrators from 28 private foundations, 35 scholars of higher education, and more than 400 randomly selected presidents of two- and four-year, public and private institutions in the United States. Using factor analysis, they identified the following categories of performance characteristics: 1) management style, 2) human relations, 3) personal image, 4) social reference, and 5) self-confidence. Based on the findings from this research, they concluded that the effective presidents are: 1) less collegial and more distant, 2) less likely to be spontaneous in speech and actions, 3) less restricted by organizational structure or by consensus of those to be led, 4) less likely to appear to make decisions easily, 5) more confident, 6) more inclined to rely on gaining respect rather than being liked, 7) more inclined to take calculated risks, 8) more committed to an ideal or vision rather than an institution, 9) more inclined to work long hours, 10) more supportive of the controversial concept of merit pay, 11) more interested in encouraging people to think differently or creatively, and 12) more likely to be concerned about higher education than with one institution (p. 111). These researchers provided correlational, factor analysis, and reliability data, to gain insights into the constructs under study; however, they did not provide convincing evidence of data validity.

In the 2000 time frame, Michael, Schwartz, and Balraj (2001) conducted a survey on indicators of president effectiveness that involved 489 trustees from universities and colleges in Ohio. The hypothesized indicators of success fell under the following

categories: 1) knowledge (of higher education, institutional politics, and differences between higher education organizations and other organizations), 2) influence (on the public, institution, politicians, fundraising, and visibility in the institution), 3) relationships (with the trustees, board chair, faculty, and students), and 4) management and leadership (academic leadership, long-range planning, budgeting, and overall institutional management). The study also examined gender, years of experience, and level of education as moderating factors. The results of the study revealed that the stronger indicators of president success as perceived by trustees are: 1) maintaining a strong relationship with the chairperson and members of the board of trustees, 2) maintaining a high level of influence within the institution, 3) managing the institution overall, and 4) supporting long range planning. Since the researchers did not provide correlational, factor analysis, or reliability data, it is impossible to assess the quality of the constructs. It is also impossible to assess the meaningfulness of the scores because the research did not provide any evidence of validity.

Nason (1997), Rhodes (2001), and others believe president performance and institutional goals are intimately related. Based on the perceived relationship between president performance and institutional outcomes, many authors (AGB, 2006; Ingram & Weary, 2000; Nason, 1997; Rhodes, 2001; Sanaghan, Goldstein, & Gavel, 2008; Schwartz, 2001) recommend the universities include criteria that address the achievement of goals and objectives that have been agreed upon ahead of time by the president and the board of regents. In some instances, goals and objectives are the only criteria applied to president performance assessment. Examples of measures associated with university goals and objectives (Trow, 1998; UNM 2007b, 2008; US News & World Report, 2008)

include: 1) student quality, 2) student success, 3) research and scholarship, 4) faculty credentials, 5) faculty productivity, 6) quality of life, 7) fundraising, 8) cultural diversity, 9) connections with the community, and 10) infrastructure modernization. Some authors (Bowen & Shapiro, 1998; Zemsky, Wegner, & Massey, 2005) emphasize that another important measure (or set of measures) that should be incorporated into performance assessment in higher education institutions is the quality of instruction. However, these authors suggest that further research is required to identify the criteria and scales that produce reliable and valid ratings in area of instruction quality.

Format

The dictionary defines *format* as “general arrangement or plan (Agnes & Guralnik, 2001, p. 556). In the context of this study, format pertains to 1) formality including documented policies, procedures, assessment instrument, and performance report, 2) purpose of the assessment, 3) structure of the assessment instrument (e.g., closed questions, open questions, objective criteria, subjective criteria), 4) participants in the assessment, and 5) frequency at which the assessment instrument should be applied. Based on research of literature and a search of the Internet, the format of the assessment instruments varies between universities.

Formality. Many higher education institutions conduct formal assessments that include written policies and written reports with the results being used for discussions or decisions on conditions of employment (e.g., salary adjustments, bonus pay, and contract renewal) (Schwartz, 1998). Some institutions use informal assessments that consist of undocumented feedback (Schuman, personal communication, November 7, 2007). Based on his experience, Schuman, a former president of two public universities, believes that

formal assessment is a rule rather than an exception in public universities and that private universities tend to conduct informal assessments. Nason (1997) reports that survey results from 1976 revealed that 49% of 116 public institutions in the sample reported they conducted formal evaluations, versus 36% that conducted informal evaluations, and 15% that did not conduct evaluations at all. This survey also showed that many governing boards that conducted informal evaluations intended to develop procedures that are more formal in the near future. A more recent survey conducted in 1997 (Schwartz, 1998) indicated that 66% of public higher education institutions had written policy requiring assessment, 81% conducted face-to-face meetings to discuss assessment results, 86% involved the reporting of assessment results to the governing board, and over 50% used president reviews for discussions or decisions on specified conditions of employment. While the 1997 survey did not specifically address the question of formality, the practices of having written policy, conducting performance reviews, and having specified conditions of employment may be indicators of a more systematic and formal assessment process. Nason (1997) believes there are several advantages of formal assessment such as providing a mechanism to: 1) focus attention on the governance structure of the institution and taking into account the attitudes, prerogatives, and behaviors of key constituent/stakeholder groups, 2) reset institutional goals or objectives after formally assessing where the institution has been and where it should be going, 3) provide rational, orderly, and systematic process with clearly articulated criteria upon which judgments can be made, 4) strengthen the president's position by revealing the complexities of the job, increasing emphasis on longer term goals rather than isolated incidents, 5) strengthen the governing board's position by providing opportunities for trustees to explore

leadership obligations, learn about constraints that impact president performance, and consider incentive available to improve performance, and 6) increase accountability of the president.

Despite the moderate use of formal assessment procedures and their perceived advantages, authors note that some people argue that formal evaluations can lead to negative consequences including a reduction in tenure of the president (Davis, W. E, & Davis, D. R., 1999; Kauffman, 1978; Nason, 1997). To go one step further, Kaufman (1978) reports that presidents of some institutions are against formal assessment because it gives faculty, students, and staff reason to believe they control the leader. Kaufman notes that some believe president assessments are diminishing, polarizing, and in the end, demoralizing to the entire community. Nason (1997) provides additional arguments that have been made against formal assessment: 1) it undermines the authority and status of the office of the president, 2) it politicizes the role of the president by subjecting them to criticism of groups that are opposed to change and innovation, 3) it is an evasion of the board's responsibility to continuously monitor performance, and 4) it offers a stage for confrontation between various constituents/stakeholders in the institution.

Purpose. The purpose of performance assessment of university presidents is similar to that of CEOs in business. Much the same as CEOs of other types of organizations, university presidents are accountable for accomplishing goals and objectives at the organization level (ACE, 2007; AGB, 2006; Barker, 2006; Clark, 2002; Fine, 1986; Hearn, 2006; Honeyman, 2007; June, 2006; Rhodes, 2001). Nason (1997) believes that the core purpose of presidential assessment is to improve his or her performance in office, and in broader terms, to improve the institution.

As mentioned previously, some universities use the results of an annual performance review for compensation setting decisions (Ingram & Weary, 2000; Schwartz, 1998; Schwartz, 2001). This compensation may be in the form of an annual salary increase and/or performance bonus (Arizona Board of Regents, 2005; Florida State University Board of Trustees, 2004). Salary increases based on merit and evaluation (expectations of performance, when and how evaluation will occur) may be specified in the president's conditions of employment (Center for Policy Analysis, 2007).

Performance assessment also serves as a means to provide feedback to the president on progress being made in the achievement of goals and objectives agreed upon between the president and the board (Nason, 1997; Schwartz, 1998, 2001). Some president employment contracts imply that performance assessments can support termination decisions by the board as evidence of substantial neglect of assigned duties and personal conduct (Arizona Board of Regents, 2005; Florida State University, 2004). Nason (1997, pp. 9-12) provides a comprehensive summary of the major purposes of assessment: 1) to fulfill the board's responsibility, 2) to strengthen the president's position and improve performance, 3) to review and improve governance of the institution, 4) to review and reset institutional goals, 5) to educate trustees, faculty, and others on the president's role, 6) to decide whether to retain or fire, 7) to set an example for faculty and staff evaluations, and 8) to set salary.

Structure. Some universities use instruments that have a series of closed-ended questions in which the raters indicate the trait and behavioral statements that most closely describe the president (Nason, 1997). Instruments with closed-ended questions frequently include numerical scales to assign ratings to each item included in the

instrument. Assessment forms typically include space to provide amplifying information or additional comments. Less frequently, the assessment forms contain open-ended questions for raters to insert subjective comments on key areas of interest such as actual achievements, performance strengths, and areas of improvement (Nason, 1997).

Rather than using a more traditional instrument with closed-ended questions and rating scales, Ingram and Weary (2000) recommend that the instrument consist primarily of open-ended questions without rating scales to “elicit comments and examples” (p. 19). They warn that assigning numerical values to leadership characteristics “widely misses the mark” (p. 19). In fact, they suggest that rating scales “trivialize the academic presidency” and disagree with the notion that “leadership assessment can be made scientific if it can be reduced to a series of numbers” (p. 19). Ingram and Weary do acknowledge that a well-constructed survey instrument can be helpful in president assessment when properly used for constructive feedback.

Participants. There is a wide range of stakeholders who may be involved in assessing the performance of a university president. Table 1 contains a list of the primary, secondary, and tertiary stakeholders along with their roles and interests in the results of president assessment (Ingram & Weary, 2000; Nason, 1997; Russ-Eft & Preskill, 2001; Schwartz, 1998). The stakeholders consist of individuals and groups, both internal and external to the institution, who have a stake in the president’s performance and who may benefit from the results of the assessment. Universities can call upon various stakeholders to serve as participants in president assessment.

Primary stakeholders. Primary stakeholders are those individuals who are typically responsible for the successful design, development, implementation, evaluation,

modification, recordkeeping, and sponsorship of a president assessment system and serve usually serve as raters.

Secondary Stakeholders. Secondary stakeholders are those groups whose representatives sometimes serve as raters and those groups directly affected by the performance of the president.

Tertiary Stakeholders. Tertiary stakeholders are those groups whose members have less frequent contact with the president, do not typically serve as raters, but have interest in the results of presidential assessment.

Research by Schwartz (1998) found that institutions solicit performance assessment inputs from the following individuals or groups: 1) trustees, 2) executive cabinet, 3) faculty 4) professional staff, 5) students, 6) president's staff, 7) foundation trustees, 8) nonprofessional staff, 9) alumni, 10) advisory board members, 11) donors, 12) government officials, 13) other presidents and 14) others. Based on Schwartz's 1997 survey of 1,348 participants, the following percentage of presidents indicated that inputs for their most recent reviews came from the following sources: 1) trustees (68%), 2) executive cabinet (32%), 3) faculty (32%), 4) professional staff (19%), 5) students (18%), 6) presidents' staff (16%), foundation trustees (16%), nonprofessional staff (16%), alumni (13%), and others (12%). Since these percentages exceed 100%, the survey results indicate that some presidents solicit feedback from multiple stakeholders for their performance reviews.

Table 1. Stakeholders

	Roles	Interests
Primary		
Board members	University governance, accountability to the governor, president selection, and oversight	Assessment of the president, fiduciary responsibility, reputation, goal achievement
President	Operations and management, fundraising, finances, and staffing	Performance assessment, status, reputation, personal rewards, and recognition
Secondary		
Administrators	Operations and management	Quality of staff and programs
Assessment committee	Organization, execution, and assessment results reporting	Quality of staff and programs, status, reputation
Faculty (including committees, deans, and department chairs)	Instruction, research, and scholarship	Quality of instructors, academic programs, academic freedom, shared governance
Students	Service consumption and program contributions	Support of personal and professional goals, ambitions
Tertiary		
Athletic associations	Intercollegiate athletics, morale, welfare and recreation	Personal and program development, competition, reputation, financial interests
Businesses	University community, job and collaborative opportunities	Social responsibility, financial interests, staff sourcing, economic development
Charitable organizations	University community	Moral/social responsibility, financial interests, staff and volunteer source
Donors	Funding source and advocacy	Philanthropy, reputation, prestige
Educational institutions and organizations	Higher education	Professional affiliations and lessons learned
Federal, state, and local government officials	Policy and funding	Accountability and effective stewardship, economic development
Financial institutions	Service to university community	Social responsibility, financial interests, staff source
General public	University programs and sponsor students	Quality of education, contributions to the community
K-12 school system	Potential university students, staff, and faculty	Higher education and employment opportunities
Media	Public information and persuasion	Social responsibility, financial interests, reputation
Parents and families	Funding source and advocacy	Safe environment, reputation, employment
Prospective administrators, staff, and students	University leadership and support positions	Promotion, opportunity for advancement, quality of life
Religious organizations	Individual development and service to the university	Moral/social responsibility, financial interests, staff and volunteer sourcing
University groups (e.g., alumni)	Individual development and service to the university	Support of personal and professional goals, ambitions

Many university boards ask presidents to submit a self-assessment on a periodic basis (Schwartz, 1998). Self-assessments may be part of the periodic assessment interview, a written self-assessment, or daily feedback (Schwartz, 1998). Several authors (Grote, 2002; Jacobs, 1986; Latham & Wexley, 1994; Mohrman, Resnick-West, & Lawler, 1989) describe the several advantages of self-assessment including its contribution to self-awareness and its utility in providing raters another important viewpoint on the ratee's performance.

Frequency. The frequency of assessment, along with the type of stakeholder who serves as a rater, are important considerations in determining the content and format of the assessment instrument. For example, stakeholders who work directly with the president on a regular basis (i.e., primary and secondary stakeholders) should be capable of completing a detailed assessment form on a more frequent basis compared to stakeholders who have limited direct contact with the president (i.e., tertiary stakeholders). Considering these factors, studies suggest that primary stakeholders be involved in annual assessments and representatives from secondary and tertiary stakeholder groups be involved in comprehensive assessments that are conducted on a less frequent basis (e.g., 3-4 or 5-7 years) (AGB, 2006; Ingram & Weary, 2000; Schwartz, 2001).

Documentation published by the AGB suggests that the instrument for an annual assessment have specific criteria compared to an assessment conducted on a 3-5 year basis that would be more general in nature. It also suggests that there be fewer participants in an annual assessment compared to the less frequent comprehensive assessment. Ingram and Weary (2000) also recommend that the president and board

participate in a joint review process conducted by external consultants at least every 5-10 years.

Methodology, Research Design, and Methods

Choosing appropriate methodology, research design, and methods are key to developing a reliable and valid performance assessment instrument. Since this study focuses on an approach for developing an assessment instrument, this literature review includes general information on research options that are available to collect and analyze relevant data. Chapter 3 provides specific details on the methodology, research design, and methods chosen for this study.

Strauss and Corbin (1998) define *methodology* as “a way of thinking about and studying social reality” (p. 3). Others suggest that it refers to the process that researchers use to design and conduct research based on worldviews (Creswell & Plano-Clark, 2007) or paradigms (Kuhn, 1996) such as postpositivism, constructivism, and pragmatism (Creswell & Plano-Clark, 2007; van Manen, 1990). The assumptions of postpositivism are: 1) there is a singular reality, 2) the researchers distance themselves from the participants, and 3) the perspectives of researchers are unbiased (Creswell & Plano-Clark, 2007). A deductive (top-down) methodology, in which researchers test a hypothesis, aligns with the assumption of postpositivism (Creswell & Plano-Clark, 2007). Research based on the worldview of postpositivism frequently employs quantitative research methods (Creswell & Plano-Clark, 2007).

The assumptions of constructivism are: 1) there are multiple realities, 2) researchers closely engage the participants, and 3) the perspectives of researchers are biased (Creswell & Plano-Clark, 2007). An inductive (bottom-up) methodology in which

researchers develop theories or hypotheses based on the views of study participants relates to the assumption of constructivism (Creswell & Plano-Clark, 2007). Research based on the worldview of constructivism often employs qualitative research methods (Creswell & Plano-Clark, 2007).

The assumptions of pragmatism are a combination of those described for postpositivism and constructivism (Creswell & Plano-Clark, 2007). The methodology in which researchers combine deductive and inductive methodologies aligns with the assumption of pragmatism (Creswell & Plano-Clark, 2007). Researchers who adopt the assumptions of pragmatism typically apply mixed-methods research using quantitative and qualitative methods (Creswell & Plano-Clark, 2007).

Research design refers to the plan of action that links the methodology to the specific research methods (Creswell, 2003; Crotty, 1998). Experimental research, quasi-experimental research, survey research, grounded theory, and mixed-methods are examples of research designs (Creswell & Plano-Clark, 2007; Fraenkel & Wallen, 2006; Shadish, Cook, & Campbell, 2002). One objective of research design is to obtain evidence that supports and refutes proposed facts and theories; to minimize measurement error; and to reduce non-response of research subjects (Fraenkel & Wallen, 2006; Marley, personal communication, April 28, 2008). *Methods* are the processes, procedures, tools, and techniques associated with data collection and analysis, such as an interview or survey (Creswell, 2003; Strauss & Corbin, 1998; van Manen, 1990). A mixed-methods design takes advantage of the multiple viewpoints and corresponds to the philosophical assumption of pragmatism (Creswell & Plano Clark, 2007; Fraenkel & Wallen, 2006; Johnson & Onwuegbuzie, 2004). Pragmatism maintains that “true ideas are those that we

can assimilate, validate, corroborate, and verify [and] false ideas are those that we cannot” (Sahakian, W. S., & Sahakian, M. L., 2005, p. 26).

The common characteristic of an experiment is to control treatments of some kind to determine their effects (Shadish, Cook, & Campbell, 2002) with the central goal to test hypotheses about cause-and-effect relationships (Fraenkel & Wallen, 2006). The goal of survey research is to obtain opinions from people on their views of a certain issue or topic (Fraenkel & Wallen, 2006). Correlational research is non-experimental or observational research in which the goal is to determine the size and direction of relationships among variables (Shadish, Cook, & Campbell, 2002). Fraenkel and Wallen (2006) suggest, “It is not uncommon for researchers to examine the relationships of responses to one question in a survey to another, or of a score based on one set of survey questions to the score based on another set” (p. 399). In this example, the investigators combine survey research and correlational research to gain insights into particular issues or questions.

Grounded theory involves the collection and analysis using a *constant comparative method* in which the researcher collects and analyzes data in an iterative process to develop generalizations of that data (Fraenkel & Wallen, 2006). Sometimes researchers develop hypotheses as qualitative generalizations, and test these hypotheses using quantitative designs such as experimental and correlational research.

Among the factors that researchers should take into consideration when choosing the research design are the goals of the study and the amount of resources (e.g., time, money, and personnel) available to complete the study (Creswell & Plano Clark, 2007). Another consideration when choosing a research design is the purpose of the research – exploration or explanation. Exploratory research tends to be qualitative in nature and

researchers often implement it within a constructivist paradigm (Creswell & Plano-Clark, 2007), but may also incorporate quantitative methods.

Grounded theory (Charmaz, 2006; Strauss & Corbin, 1990) is an example of a qualitative method and exploratory factor analysis (EFA) (Pett, Lackey, & Sullivan, 2003) is an example of a quantitative method that applies to exploration. Exploratory research (Fraenkel & Wallen, 2006; Stevens, 2002) is useful in identifying the tasks (i.e., dimensions) and categories of tasks (i.e., constructs) that an individual may perform in their job and stakeholder perceptions of the relative importance of performing these tasks. These constructs and dimensions provide a basis for the assessment criteria, measures, and standards of performance. Researchers can use assessment criteria (i.e., the set of desirable traits, behaviors, and performance outcomes), the measures, and the standards of performance to develop the items (questions), ratings scales, and format of a questionnaire or survey for performance assessment. Exploratory research is also useful for generating hypotheses that can be tested later using explanatory (confirmatory) research methods (Charmaz, 2006; Pett, Lackey, & Sullivan, 2003).

Explanatory (confirmatory) research is useful in determining relationships among variables (Fraenkel & Wallen, 2006; Stevens, 2002) such as assessment criteria. For example, a researcher can use correlational research to determine the importance of various candidate criteria for a performance assessment instrument. Multiple regression is a correlational research method that enables a researcher to determine what variables (e.g., traits and behaviors) are the best predictors of criterion variables (e.g., performance outcomes) (Fraenkel & Wallen, 2006). Confirmatory factor analysis (CFA) is another correlational research method that researchers use to test hypotheses concerning the

relationship between variables (Pett, Lackey, & Sullivan, 2003). Sometimes researchers use CFA to confirm the relationships among variables as a follow-up to EFA.

Other statistical analysis methods that researchers sometimes apply for explanatory research are analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) (Tabachnick & Fidell, 2007). Researchers use ANOVA and MANOVA to investigate how groups differ on a variable or sets of variables, respectively. For example, researchers can use MANOVA to identify how various groups differ on their perceptions of what assessment criteria should be included in an assessment instrument for an individual. MANOVA may also be useful in obtaining validity evidence by determining if there are significant differences between the scores from an assessment performed by one group versus another group. If there is not sufficient evidence that the assumptions for multiple regression, ANOVA, and MANOVA are tenable, researchers can use non-parametric statistics such as Mann-Whitney and Friedman One-Way ANOVA tests (Field, 2005; Fraenkel & Wallen, 2006) as alternative methods, however, these tests usually result in the loss of power (i.e., the ability to identify differences when they truly exist).

Common methods to collect data for exploratory and explanatory research include sampling, questionnaires (e.g., surveys), individual interviews, focus group interviews, and document analysis (Crotty, 1998). Researchers use sampling procedures to decide the best people and sites from which to collect data to address research questions (Creswell & Plano-Clark, 2007). For *purposive sampling*, researchers select participants who have experience with the issues or questions of interest (Creswell & Plano-Clark, 2007). On the other hand, *probabilistic (random) sampling* involves the random

selection of participants by using a tool such as a random number generator or table (Creswell & Plano-Clark, 2007). Questionnaires (surveys) are a common instrument to collect data for exploratory and explanatory designs. Researchers employ measurement, scaling, and item analysis techniques to determine the content and format of questionnaires and surveys (Creswell & Plano-Clark, 2007; Nunnally, 1978) based on the constructs and dimensions of interest. When researchers want to collect data from a selected participant or relatively small group of participants, they use individual or focus group interviews, respectively (Creswell & Plano-Clark, 2007). Researchers may use document analysis during a literature review to identify a theory to test or questions they should ask in surveys and interviews (Creswell & Plano-Clark, 2007).

Researchers can use pilot surveys to gain insights into the variability of data and to determine the sample size necessary to achieve the desired level of power for statistical tests (Rea & Parker, 2005; Tabachnick & Fidell, 2007). *Power* is the probability that a test will reveal statistically significant differences between variables if they truly exist (Cohen, 1988, 1992). *Statistical Power Analysis for the Behavioral Science* (Cohen, 1988) is an authoritative source of information for determining the required sample size of a survey. Lenth (2007) provides a tool referred to as Piface[®] that researchers can use to calculate sample size requirements for various statistical tests as well.

In addition to using pilot survey data to determine the desired sample size for a statistical test, researchers consult with subject matter experts (SMEs) to determine the content and format of the final survey instrument (Fraenkel & Wallen, 2006). Another advantage of a pilot survey is that it serves as a means for researchers to rehearse the entire data collection and analysis process. A pilot survey may also provide early insight

into the tenability of the assumptions associated with statistical tests as an early step in determining data reliability and validity even though the researchers do not use pilot survey data for quantitative analysis.

In mixed-methods research designs, researchers sometimes use the qualitative results from the pilot survey, individual interviews, and focus group interviews to design the instrument that they will use to collect data for quantitative analysis (Creswell & Plano-Clark, 2007). When completing quantitative analysis, researchers occasionally use qualitative data from follow-up individual interviews and focus group interviews to explain the results of quantitative analysis (Creswell & Plano-Clark, 2007). For example, qualitative analysis may reinforce the evidence of important relationships between variables discovered during quantitative analysis and may help researchers substantiate cause and effect relationships between these variables.

Creswell and Plano-Clark (2007) advocate that researchers use an exploratory design to develop survey instruments they will use to collect quantitative data. Creswell and Plano-Clark's instrument development model consists of the following steps:

- 1) collect qualitative data, 2) analyze qualitative data, 3) determine qualitative results,
- 4) develop quantitative data collection instrument with appropriate items and scales,
- 5) collect quantitative data, 6) analyze quantitative data, 7) determine quantitative results,
- 8) interpret quantitative results using qualitative results to validate the quantitative results.

In their instrument development model, Creswell and Plano-Clark connect the qualitative and quantitative methods through the development of the instrument and the validation of the qualitative and quantitative data. Applying the Creswell and Plano-Clark model, researchers systematically compare the data from all qualitative and

quantitative data sources through the process of *triangulation* to obtain evidence of the reliability and validity of data.

One of the key aspects of research in general and individual performance assessment in particular is the psychometric properties of the instruments that researchers and raters use to collect data and summarize the scores. Psychometric properties of an instrument include evidence of reliability and validity (Wilkinson, 1999). Wilkinson warns that researchers sometimes give insufficient attention to the quality of their instruments. He goes on to stress that researchers should pay special attention to the psychometric properties of their instruments “to prevent the accumulation of results based on unreliable or invalid measures” (Wilkinson, 1999, p. 596).

A prerequisite for having *valid* data is having *reliable* data. *Reliability* is the “consistency of repeated measurements across persons” (Carmines & Zeller, 1979, p. 31). Fraenkel and Wallen (2006) expand on this definition: “Reliability refers to the consistency of scores – how consistent they are for each individual from one administration to another and from one set of items to another” (p. 157). Referring to *classical test theory*, the observed score is the sum of true score plus random error, and if the measurement across persons contains excessive random error, the score is unreliable (Carmines & Zeller, 1979). The American Educational Research Association, American Psychology Association, and National Council on Measurement in Education (2004); Carmines and Zeller (1979); Cronbach (1951); Fraenkel and Wallen (2006); and Latham and Wexley (1994) describe several techniques for determining the reliability of quantitative data.

The *retest* technique involves administering the same data collection instrument to the same participants after a certain period and then comparing results for consistency (Cronbach, 1951). *Test-retest* is another name for retest (American Educational Research Association, American Psychology Association, & National Council on Measurement in Education, 2004). The *alternate forms* technique involves giving a similar but different instrument to the same participants and then comparing results (Fraenkel & Wallen, 2006). Additional terms for alternate forms are *equivalent forms* and *parallel forms* (Fraenkel & Wallen, 2006).

The *split-halves* technique allows the researcher to check data reliability by dividing a data into two sets and then correlating the scores (Carmines & Zeller, 1979). Split-halves is an *internal consistency* technique in that it requires only a single administration of an instrument to estimate reliability (Fraenkel & Wallen, 2006). Latham and Wexley (1994) define internal consistency in practical terms as “an indication of the homogeneity, or ‘sameness,’ of the items that comprise a scale” (p.68) in an assessment instrument. One of the more commonly used tests for internal consistency is *Cronbach’s alpha* because of its ease of use and its incorporation of positive aspects of other reliability techniques (Carmines & Zeller, 1979). The *Kuder-Richardson* approaches (*KR-20* and *KR-21*) are additional internal consistency techniques. *KR-20* is a special case of Cronbach’s alpha that researchers use for analysis involving dichotomous variables (e.g., agree/disagree, for/against, and yes/no) (Fraenkel & Wallen, 2006). *KR-21* is the simplest calculation of internal consistency that only requires the number of items in the instrument, the mean of the scores, and the standard deviation of the scores (Fraenkel & Wallen, 2006).

Another key, but often downplayed, aspect of research and measurements that is associated with reliability is *validity*. There are several definitions for validity. One of the more traditional definitions of validity is the “extent to which a measure measures what it purports” (Carmines & Zeller, 1979, p. 17). Another definition of validity is “the overall degree of justification for test interpretation and use” (Messick, 1980, p. 1012). One of the more contemporary definitions of validity is “the truth of, correctness of, or degree of support for an inference” (Shadish, Cook, & Campbell, 2002, p. 513). Bernardin and Beatty (1984) believe that “validity is the ‘sine qua non’ not just of performance appraisal but of any assessment procedure” (p. 143). They go on to say, “validity in the context of performance appraisal is the extent to which the ratings on an appraisal instrument correspond to the actual levels for those who are rated” (p. 143).

There are seven types of validity associated with quantitative analysis:

1) statistical conclusion validity, 2) internal validity, 3) content validity, 4) construct validity, 5) external validity, 6) criterion validity, and 7) consequential validity (American Educational Research Association, American Psychology Association, & National Council on Measurement in Education, 2004; Fraenkel & Wallen, 2006; Latham & Wexley, 1994; Lawshe, 1975; Messick, 1980; Nunnally & Bernstein, 1994; Shadish, Cook, & Campbell, 2002; Shepherd, 1993). Evidence of *statistical conclusion validity* includes compliance with assumptions of the statistical tests, control of Type I error, and availability of sufficient power to identify statistically significant results (Shadish, Cook, & Campbell, 2002). Type I errors occur in statistical testing when the researcher incorrectly rejects the null hypothesis concluding there is a difference between groups or variables when there is not.

Evidence that there is a defensible cause and effect relationship between two variables not explained by other factors supports *internal validity* (Shadish, Cook, & Campbell, 2002). Fraenkel and Wallen (2006) contend that internal validity requires that “any relationships observed between two variables must be unambiguous as to what it means rather than being due to something else” (p. 169). Examples of “something else” include differences in age, gender, ethnicity, and experience as well as the loss of research subjects, and researcher biases (Fraenkel & Wallen, 2006).

Another form of validity evidence that is particularly important for individual performance assessment is *content validity*. Content validity requires accurate and complete descriptions of the essential elements that comprise the domain of interest (Carmines & Zeller, 1979; Lawshe, 1975). According to Nunnally and Bernstein (1994), the two standards for ensuring content validity are: 1) the instrument should have a representative collection of items and 2) the instrument should be constructed using sensible methods. For performance assessment, the essential elements (i.e., items in the assessment instrument) may include the more important traits, behaviors, and performance outcomes attributable to the ratee (Latham & Wexley, 1994).

Construct validity relates to the proper *operationalization* of the domain of interest (Shadish, Cook, and Campbell, 2002). Operationalization refers to the methods used to represent a construct such as ensuring the proper alignment of the construct and its dimensions within a domain of interest (Shadish, Cook, and Campbell, 2002). For example, the construct of leadership may include traits of “assertiveness” and “agreeableness” within the domain of president performance. These traits are the dimensions of the construct of leadership. Having the scores for the variables of

assertiveness and agreeableness that correlate highly on a survey provides evidence of *convergent validity* (American Educational Research Association, American Psychology Association, & National Council on Measurement in Education, 2004; Campbell & Fiske, 1959; Shadish, Cook, and Campbell, 2002) as part of construct validity. The construct leadership should not include a dimension (i.e., trait) such as “demonstrates fiscal control” since this should more closely align with the construct of management rather than with leadership. The lack of correlation between variables that theoretically should align with different constructs is evidence of *discriminant validity* (American Educational Research Association, American Psychology Association, & National Council on Measurement in Education, 2004; Campbell & Fiske, 1959; Shadish, Cook, and Campbell, 2002), which is another aspect of construct validity.

Operationalization of the domain of interest also includes the process of developing appropriate measures for the variables that represent the constructs and their corresponding dimensions (Campbell & Fiske, 1959; Shadish, Cook, and Campbell, 2002). Additional evidence of construct validity is gained by confirming theoretically predicted associations of measures of other constructs (Campbell & Fiske, 1959). For example, a researcher can obtain evidence of construct validity by determining there is a statistically significant relationship between various leadership traits and various performance outcomes based on hypotheses derived from theory.

Nunnally and Bernstein (1994) emphasize, “The degree to which it is necessary and difficult to validate measures of psychological variables is proportional to the degree to which those variables are concrete or abstract” (p. 84). Accordingly, researchers can save themselves time and effort in the validation process by defining measures that are

understandable and representative of the construct of interest. Shepherd (1993) stresses that researchers can make constructs more understandable by using internal and external models. An internal model shows the relationships of the dimensions within a given construct. An external model illustrates the relationships between different constructs that pertain to the domain of the study.

Evidence of *external validity* relates to the generalizability of findings to other populations, settings, and conditions (Fraenkel & Wallen, 2006). *Population generalizability* is the degree to which a sample represents the target population of interest (Fraenkel & Wallen, 2006). *Ecological generalizability* refers to the degree to which a researcher can extend the results of a study to different settings and conditions (Fraenkel & Wallen, 2006). Using random selection of survey participants and having a favorable response rate to surveys contributes to external validity (Shadish, Cook, and Campbell, 2002).

Researchers obtain evidence of *criterion validity* by comparing the scores obtained from one instrument (the one under evaluation) to the scores on a second test or procedure (referred to as the criterion) that is presumed to measure the same variable (Fraenkel & Wallen, 2006). For example, if the leadership ratings a president receives on an annual performance assessment correlate highly with the achievement of university goals (i.e., criteria) related to fundraising, stakeholder satisfaction, and student-faculty ratio, some researchers would consider this evidence of criterion validity. There are two forms of criterion validity – *predictive validity and concurrent validity* (Fraenkel & Wallen, 2006). Comparing the ratings a president receives at the end of the spring semester to the achievement of goals at the end the calendar year will provide evidence of

predictive validity. Comparing the ratings a president receives from members of the board of regents to the ratings received from members of the faculty will provide evidence of concurrent validity.

Consequential validity relates to the appropriateness of making value implications based on study results and the social consequences of disseminating these study results (American Educational Research Association, American Psychology Association, & National Council on Measurement in Education, 2004; Fraenkel & Wallen, 2006; Messick, 1980). While the concept of consequential validity is a controversial subject (Fraenkel & Wallen, 2006) and is not particularly relevant to this study, it may be a consideration when performing an actual performance assessment of university presidents. For example, lower scores on an annual assessment of a president may result from deficiencies in the assessment process or instrument rather than from poor performance of the president. In this case, evidence that supports consequential validity would include verifying the limitations of the assessment and taking steps to preclude unintended consequences for the president such as loss of motivation, loss of reputation, and reduction in monetary rewards (e.g., bonus or merit pay). Consequential validity may also come into play when considering the release of the results of an actual performance assessment of a university president because of potential legal ramifications.

Reliability and validity apply to the trustworthiness of qualitative data as well (Guba & Lincoln, 1981). In qualitative research, the trustworthiness of data relates to its *dependability, credibility, transferability, and confirmability* (Creswell, 2006; Guba & Lincoln, 1981; Lincoln & Guba, 1985). *Dependability* is the qualitative equivalent of reliability (Lincoln & Guba, 1985). Researchers obtain evidence of dependability by

checking the consistency of data through techniques such as triangulation (Creswell, 2006). For example, having consistent findings from individual interviews, focus group interviews, and surveys provide evidence of dependability.

Credibility is similar to internal validity (Guba & Lincoln, 1981). Credibility requires testing the truth-value of findings and interpretations with various sources. Researchers can use techniques such as *external audits* and *member checking* to increase the credibility of qualitative data. In an external audit, researchers solicit feedback from subject matter experts (SMEs) on whether or not the data supports the findings, interpretations, and conclusions of the study. These SMEs should have no connection to the study. Member checking is the use of study participants to provide feedback on findings and interpretations. Lincoln and Guba (1985) believe that member checking is “the most critical technique for establishing credibility” (p. 314).

Transferability is the qualitative term for external validity that pertains to the generalizability of quantitative data (Lincoln & Guba, 1985). In qualitative research, techniques such as writing *thick descriptions* of feedback received from study participants and comparing the demographics of the samples to those of the target population contribute to external validity (Creswell, 2006; Lincoln & Guba, 1985).

Finally, *confirmability* in qualitative research is similar to objectivity in quantitative research in that the data is factual and confirmable from other sources (Guba & Lincoln, 1981). Verifying that the findings from research stem from characteristics of the research subjects and context rather than from the biases, motivations, interests, and perspectives of the researcher is evidence of confirmability. Keeping detailed reflexive notes from interviews and surveys, and using SMEs to confirm the constructs,

dimensions, measures, and methodology are examples of techniques that support confirmability (Lincoln & Guba, 1985).

Evaluation of Previous Research

Supporting Evidence

Literature on performance assessment provides strong evidence of the importance of this process in professional and organizational development (Armstrong, 2009; Association of Governing Boards of Universities and Colleges, 2006; Bernardin & Beatty, 1984; Grote, 2002; Ingram & Weary, 2000; Kauffman, 1978; Latham & Wexley, 1994; Mohrman, Resnick-West, & Lawler, 1989; Nason, 1997; Schwartz, 1998; Sokol & Oresick, 1986). If done well, performance assessment can serve as an effective focusing and feedback mechanism for the individual and the organization. Some literature suggests that performance assessment, particularly when it is based on a sound job analysis, can also serve as useful evidence in resolving employment disputes and litigation.

In recognition of the positive aspects of this process, the current trend is in the direction of more formal and more elaborate assessments (Nason, 1997); although Schwartz (1998) found most university presidents are assessed using an informal process. The fact that a relatively high percentage of universities use the results of annual performance assessments to make decisions on president compensation (Association of Governing Boards of Universities and Colleges, 2006; Schwartz, 1998, 2001) is further evidence of the importance of president assessment. Literature (Association of Governing Boards of Universities and Colleges, 2006; Ingram & Weary, 2000; Nason, 1997) also suggests that performance assessment of presidents is becoming more

important to key stakeholders who are calling for increased accountability, transparency, and performance in educational institutions.

Researchers and practitioners can find a wealth of information on theories and concepts associated with learning, leadership, management, followership, organization, and evaluation that is applicable to the development of an assessment instrument.

Literature written by university presidents, administrators, and faculty (Alfred, 2006; Bowen & Shapiro, 1998; Brown, 2006; Bruce, 2008; Duderstadt & Womack, 2003; Keohane, 2006; Padilla, 2005; Rhodes, 2001; Sanaghan, Goldstein, & Gavel, 2008) and published by the AGB (Ingram & Weary, 2000; Nason, 1997; Schwartz, 2001) provides useful information that can be applied to develop the content and format of an assessment instrument for a university president. Valuable information on presidential assessment is available from many universities that post policies, procedures, and assessment instruments on their websites (Arizona Board of Regents, 1990, 2005; University of Alabama, 2003; University of Michigan, 2006; University of Nevada Las Vegas, 2000, 2003; University of New Mexico, 2004, 2007a, 2007b, 2008, 2009; University of North Florida, 2006; University of Utah System Board of Regents, 2006; University of Washington Board of Regents, 2006). Based on these sources, it appears that many universities continue to assess president traits and behaviors. However, it appears that universities are placing greater emphasis on the use of performance outcomes (e.g., achievement of goals and objectives) as assessment criteria with the president usually negotiating expected performance outcomes with the board of regents (Association of Governing Boards of Universities and Colleges, 2006; Ingram & Weary, 2000; Nason, 1997). This emphasis on performance outcomes is consistent with *results-oriented*

assessment approaches (Mohrman, Resnick-West, & Lawler, 1989) in which assessment ratings focus on the results achieved rather than how the results are accomplished.

Addressing the issue of the appropriate research design, methodology, and methods for this study, an exploratory design that incorporates mixed-methods allows for a systematic investigation into the domain of president performance assessment from multiple viewpoints. Involving multiple raters (such as superiors, peers, subordinates, individuals outside the organization, and the employees themselves) in the assessment process can serve as a means to increase the reliability and validity of ratings and to reduce biases that can negatively affect the ratings (Bernardin & Beatty, 1984; Latham & Wexley, 1994). Using multiple raters to provide different perspectives on a president's performance is analogous to using mixed-methods in that both approaches generate more data and allow for the triangulation of that data.

Authors acknowledge that validity is a crucial consideration, not only for performance appraisal, but for any assessment procedure (American Educational Research Association, American Psychology Association, & National Council on Measurement in Education, 2004; Bernardin & Beatty, 1984; Latham & Wexley, 1994; Nathan & Cascio, 1986; Sokol & Oresick, 1986) and that reliability is a prerequisite for validity (American Educational Research Association, American Psychology Association, & National Council on Measurement in Education, 2004; Latham & Wexley, 1994). Given the increasing competition for research funding and the increasing emphasis on accountability, transparency, and product/quality improvement, the need to provide evidence of reliability and validity in research and in performance assessment may grow as well. There is a vast amount of information available on the subjects of

reliability and validity. Researchers and organizations should take advantage of these resources in their research designs and in their performance assessment processes and tools, respectively.

Contradictory Evidence

Surveys conducted by Fisher and Tack (Fisher, Tack, and Wheeler, 1988) and Michael, Schwartz, and Balraj (2001) reveal there are differences in opinions on the characteristics of effective presidents. For example, Fisher and Tack's survey showed that presidents were more effective if they were more aloof whereas Michael, Schwartz, and Balraj's study indicated that positive working relationships with key stakeholders (i.e., trustees, the board's chair, faculty, and students) are "critical to the president's influence" (p. 344). This is just one of the many examples of authors and researchers who have identified conflicting indicators of successful university presidents or who have identified indicators not contained in other literature on the subject.

Some individuals believe that there are few, if any, traits or human qualities that can serve as normative measures of performance (Armstrong, 2009; Bernardin & Beatty, 1984; Ingram & Weary, 2000; Latham & Wexley, 1984). These individuals suggest that behaviors or performance outcomes (such as the achievement of goals) are more accurate and relevant indicators of individual performance. In opposition to this viewpoint, Coens and Jenkins (2002) argue that there is little relationship between the performance of an individual and the performance outcomes of an organization because there are so many intervening variables and interdependencies. Returning to the issue of traits, Nason (1997) insists that "any assessment of [university] presidential performance must take some account of character, personality, and style of the president" (p. 38).

Compared to literature that supports formal performance assessment, there is less literature that points out criticisms of formal assessment (Armstrong, 2009; Coens & Jenkins, 2002; Kauffman, 1978; Nason, 1997; Ingram & Weary, 2000). The authors who present an opposing view on the usefulness of performance assessment argue that the process can be demeaning and demoralizing. Some authors (Bernardin & Beatty, 1984; Coens & Jenkins, 2002; Latham & Wexley, 1994) point out that biases are inherent to the assessment process (e.g., rater errors) and can negate the reliability and validity of ratings. As an alternative to formal assessment that is usually performed on an annual basis, some authors suggest that informal feedback be given to employees as an informal, ongoing process. Coens and Jenkins (2002) suggest that organizations eliminate performance assessment altogether and replace it with coaching and mentoring.

Some literature on university president assessment suggests that performance assessment can contribute to higher turnover of presidents to the detriment of these institutions (Ingram & Weary, 2000; Kauffman, 1978). One of the adverse effects of high turnover is lack of continuity in setting and accomplishing longer term goals and objectives. Literature points out that some people believe that university presidents should be exempt from formal performance assessment with specific performance criteria and standards because of their unique status as senior executives (Ingram & Weary, 2000; Kauffman, 1978; Nason, 1997). These people imply that the complexities of the job and the importance of protecting the reputation of the president outweigh the benefits of a formal assessment.

Addressing the issue of research design, methodology, and methods, mixed-methods studies “are complex and may require extensive time, resources, and effort on

the part of the researcher” (Creswell & Plano-Clark, 2007, p. 181). In a similar vein, involving representatives from several university stakeholder groups can be a formidable undertaking. However, if the president and key constituents, such as the board of regents and faculty, believe that assessment of the president is important, and not just a pro forma or square-filling process, then they should consider investing the necessary resources to develop and maintain an effective assessment system. There is evidence that using a mixed-methods approach to develop an instrument and that using multiple raters in the process can increase the reliability and validity of performance assessment data as well as maximize its utility to the individual and the institution.

Gaps in Knowledge

A significant gap in the assessment knowledge base is the lack of a rigorous and proven approach for “end to end” development of a performance assessment instrument. This limitation exists in literature that generally addresses performance assessment as well as literature that specifically addresses assessment within universities. An end-to-end approach refers to one that begins with the consideration of appropriate theories and concepts and ends with having a viable process in place to update the assessment instrument. It should also include a mechanism for reviewing and incorporating applicable theories and concepts, stakeholder perspectives and priorities, and organization and external factors into revised versions of the assessment instrument.

Research also suggests there are gaps in knowledge of the constructs, along with the dimensions of those constructs that adequately define the domain of university president assessment. Without clearly defined constructs, it is difficult to operationalize the constructs through the development of criteria and standards that form the basis of the

items and rating scales that may be included in the performance assessment instrument. Poor operationalization of the constructs also jeopardizes the reliability and validity of the ratings obtained from an assessment instrument.

Most of the literature on individual performance assessment appears to be based on practical experience. There is a significant amount of research on leadership and management theories and concepts. This study attempts to connect relevant theories/concepts to their application in the context of performance assessment. The lack of connection between the theories and concepts associated with university president assessment and the actual process and tools used in this assessment raises the question about the reliability and validity of performance ratings derived from such a system. Unanswered questions about the reliability and validity, in turn, can cause stakeholders to lose confidence in the fairness and meaningfulness of the assessment system and to settle for a less effective and less efficient approach.

Justification for Current Study

While performance assessment is a widespread practice, literature reveals there are shortcomings that affect how well it is accepted and implemented by universities. One of the shortcomings is the lack of convincing evidence that performance assessment serves as an accurate indicator of individual performance, promotes individual improvement, and contributes to organizational performance. Another limitation is the lack of evidence of the reliability and validity of the data obtained from a performance assessment instrument.

The purpose of this study is to define an approach for developing a performance assessment instrument for university presidents that will produce reliable and valid

ratings that can be used for personal development, organizational development, and personnel decisions involving the president. This process will account for relevant theories/concepts associated with assessment in university environments. It will include steps for identifying external factors that can affect president performance ratings. Other key products from this study will be recommendations on appropriate content and format of the assessment instrument itself as well as a set of related hypotheses for testing in future research.

CHAPTER THREE

METHODOLOGY

This chapter provides details of the methodology and procedures for collecting, analyzing, and reporting the findings from this research on university president assessment as well as the process for developing an assessment instrument as introduced in Chapter 2. The contents of this chapter include the 1) theoretical and conceptual framework, 2) methodology, research design, and methods, 3) data management and analysis, 4) risk and mitigations, and 5) summary.

Theoretical and Conceptual Framework

A key finding from the literature review was the gap in knowledge of the constructs and corresponding dimensions associated university president performance. Lacking a clear understanding of these constructs and dimensions can affect the quality of instruments that universities use to assess their presidents. One of the purposes of this study was to identify a preliminary model that shows the relationship among variables derived from the literature review as a point of departure. An exploratory research approach provided information to determine if the preliminary model was accurate based on the perspective of the target population, to answer the research questions, and to set the stage for follow-up explanatory research that is beyond the scope of this dissertation. For this study, the problem statement and research questions served as a foundation for the theoretical and conceptual framework.

Problem Statement

A comprehensive approach is not available that universities can use to develop the content and format of a president assessment instrument that provides sufficient evidence of the reliability and validity of the ratings or scores derived from this instrument.

Research Questions

1. What approach can UNM and other public universities use to develop an effective performance assessment instrument for their presidents?
2. What is the preferred content and format for a president performance assessment instrument?

Hypotheses

This study incorporated exploratory research methodology to identify an effective approach for developing performance assessment instruments as well as the content and format of those instruments. Since there were many unanswered questions going into this study about the appropriate content of an assessment instrument, an exploratory research design was necessary to consolidate the very large number of items that could serve as measures (Nunnally & Bernstein, 1994) for assessing the performance of university presidents. Among the main products of this study are proposed hypotheses on the relationships between variables pertinent to president assessment.

Future investigation of the relationships between variables may provide insight into the relative importance of these variables from the perspectives of additional target populations (e.g., members of the board of regents, administrative staff, students, etc.). Having a better understanding of the relative importance of variables can help assessors make more informed decisions on instrument content and format. For example, assessors

may limit the items in an assessment instrument to those that are most important to reduce the time and effort to complete the assessment, analyze the results, and focus attention on the more important performance factors and outcomes. Chapter 5 of this document lists these hypotheses and proposes follow-up tests.

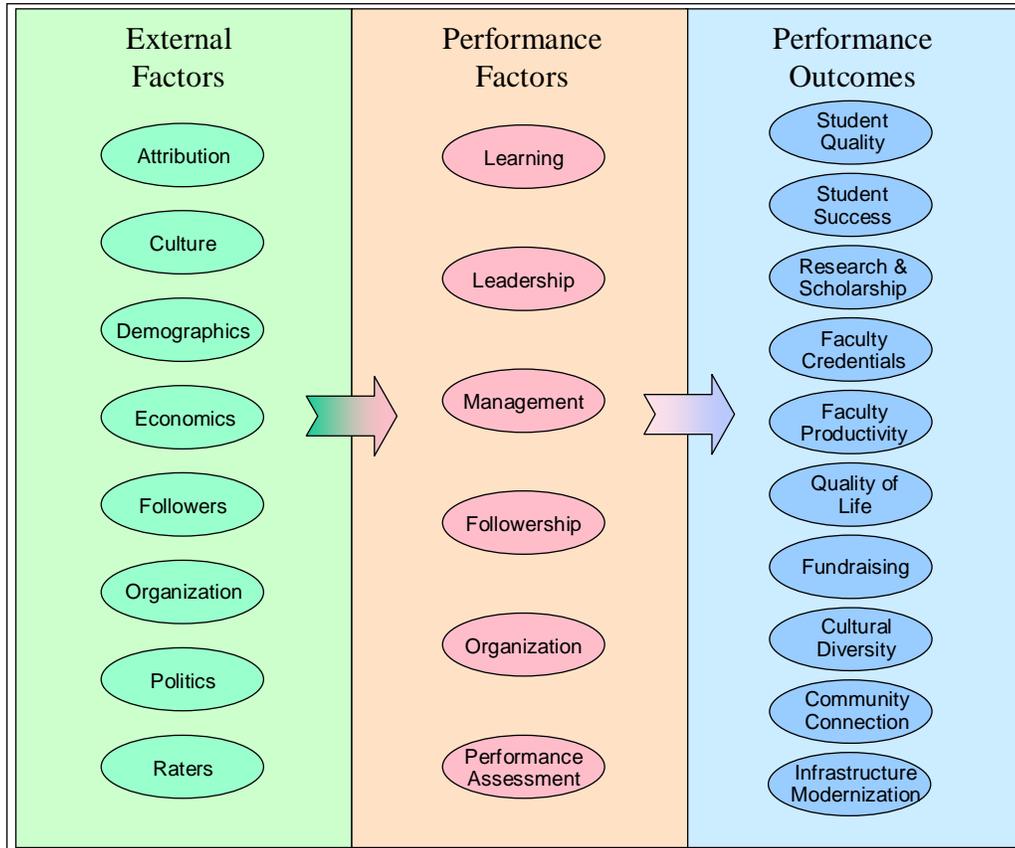


Figure 3. Preliminary Theoretical and Conceptual Framework

Figure 3 presents the preliminary external model that illustrates the relationships among the variables examined during this study. Chapter 5 contains a new model based on the findings of this dissertation research. The literature review provided evidence that various theories (e.g., learning, leadership, management, etc.) can serve as a basis for factors that reflect the performance of a university president. In turn, these factors may translate into criteria to include in an assessment instrument. Literature supported the

notion that president performance factors can influence performance outcomes such as the achievement of university goals and objectives in the areas of student quality, student success, research, and scholarship, etc. The literature review also revealed there are several external factors that can affect the perceived performance of a president and that raters should consider these factors when performing an assessment.

Methodology, Research Design, and Methods

This study incorporated the philosophical framework and fundamental assumptions associated with postpositivism, constructivism, and pragmatism (Creswell & Plano Clark, 2007). It included quantitative methods that involve a top-down, deductive reasoning approach derived from the framework and assumptions associated with postpositivism. Consistent with constructivism, this study incorporated a qualitative research design that involved a bottom-up, inductive reasoning approach. Since this study integrated postpositivist and constructivist frameworks and qualitative and quantitative methods, it aligned with pragmatism and a mixed-methods design. In addition, this study incorporated aspects of exploratory and triangulation research along with survey and correlational designs (Creswell & Plano Clark, 2007).

Rationale for Methodology, Research Design, and Methods

The reasons for choosing the methodology, research design, and methods for this study were as follows:

- The procedures incorporated multiple philosophical paradigms and methods to address the problem identified during this study and the research questions.
- The procedures provided a systematic for ensuring the trustworthiness, reliability, and validity of the data collected during the review of archival data,

individual interviews, focus group interviews, and surveys through triangulation (Russ-Eft and Preskill, 2001).

- The procedures provided multiple decision points during the study at which the researcher could make adjustments to improve the efficiency and effectiveness of the data collection and analysis process.
- The procedures included an iterative process allowing the researcher to use findings from previous steps in the study process to tailor subsequent steps in the process.
- The procedures provided a framework for creating a robust approach to development of a president assessment instrument.
- The procedures consisted of complementary steps to determine the preferred content and format of a president assessment instrument from the perspective of a key constituent that shares in the governance of a university.

Assessment Instrument Development Approach

The Assessment Instrument Development Approach (AIDA) proposed in this study is based on the literature review. AIDA embodies the methodology, research design, and methods selected for this study. Figure 4 illustrates a proposed approach for developing the content and format of a president assessment instrument. This approach includes the complementary use of quantitative and qualitative research methods to address the problem statement and to answer the research questions. AIDA provides a means to connect the theoretical and conceptual aspects of senior leader assessment to the process of determining what assessment criteria should be included in the assessment instrument and the format of the instrument.

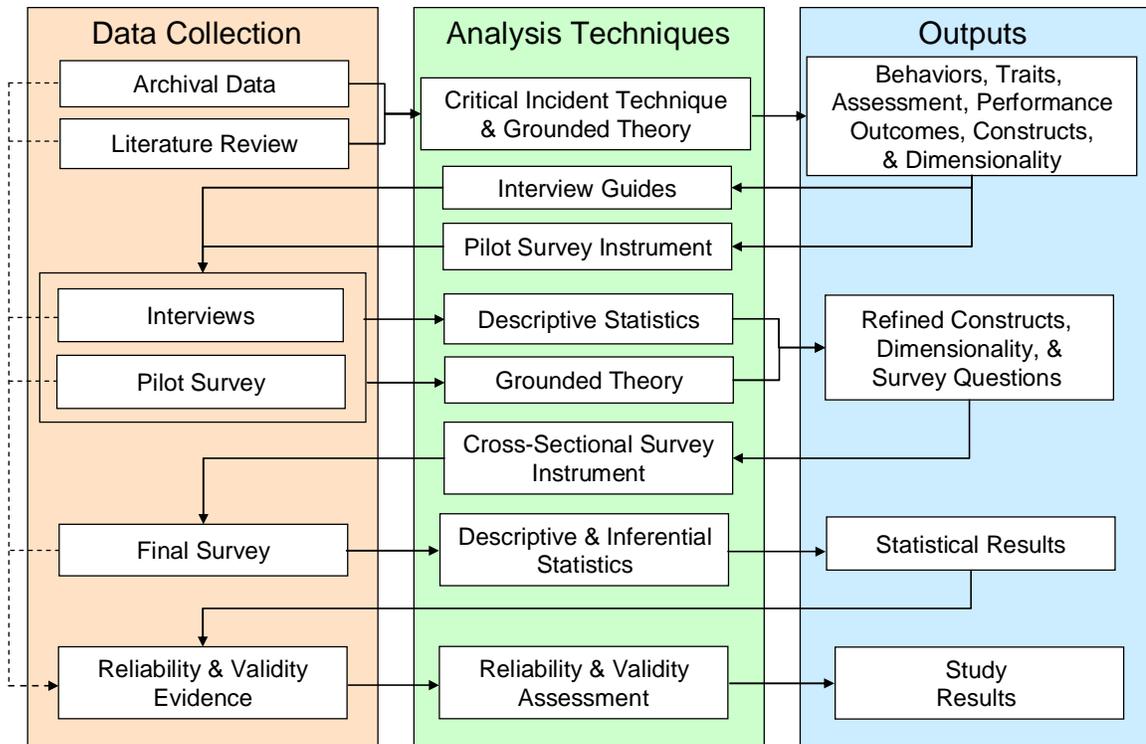


Figure 4. Assessment Instrument Development Approach

The steps in this approach fall under the major headings of Data Collection, Analysis Techniques, and Outputs. As a complement to top-down, deductive reasoning typically applied in evaluation and scientific method, AIDA also includes bottom-up, inductive reasoning to identify the content as well as the format of the assessment instrument. This chapter elaborates on the actual application of the methodology, research design, and methods to formulate the results and discussion included in chapters 4 and 5 of this document.

The first step in implementing AIDA for this study was conducting a review of archival data and literature (Fraenkel & Wallen, 2006; Russ-Eft & Preskill, 2001) that encompassed the theory, concepts, and practical application of the performance assessment and included the criteria, measures, and format of instruments used for senior

executives. Archival data included demographic information on the university used to determine if the respondents to the surveys were representative of the target population. President assessment policies, processes, procedures, and instruments from universities were additional sources of archival data that provided information about content and format of existing assessment tools.

The next step was applying critical incident technique (CIT) and grounded theory methods to identify preliminary constructs and dimensions that formed the basis for the variables, qualitative measures, and quantitative measures for the study. Applying the inductive, bottom-up approach prescribed by CIT and grounded theory, the research identified the dimensions (i.e., traits, behaviors, assessment considerations, and performance outcomes) associated with the domain of president performance assessment. The dimensions identified in this step reinforced the preliminary constructs derived from the theories described in the literature review that included: 1) president performance factors, 2) president performance outcomes, and 3) external factors that can have an impact on performance. These dimensions also served as a basis for the preliminary theoretical and conceptual framework illustrated in Figure 3.

The next AIDA step was operationalizing the constructs and dimensions developed in the previous step. The process of operationalization consisted of defining items to be included in the interview guides and pilot survey. The constructs and dimensions served as a basis for the items included in individual and focus group interview guides. Appendix B includes the individual interview guides and Appendix C includes the focus group interview guides. The pilot survey included a Likert scale to

measure the variables derived from the preliminary theoretical constructs and dimensions. See Appendix D for an example of the pilot survey.

The next step included conducting the interviews and pilot survey in parallel with each other. Individual and focus group interviews with UNM faculty members were conducted. The interview participants were faculty members who had expertise in the constructs of interest in this study: 1) learning, 2) leadership, 3) management, 4) followership, 5) organization, 6) performance assessment, 7) performance outcomes, and/or 8) external factors that can affect performance. Some of the interviewees had experience with the process of shared governance at UNM. A cross section of colleges and schools at UNM were used in the selection of faculty members for the interviews. UNM faculty members who participated in the interviews were from the Anderson School of Management, College of Art and Sciences, College of Education, School of Engineering, School of Law, and School of Medicine. StudentVoice™ administered the web-based pilot survey. The UNM Office of the President and Office of the Provost provided approval to use the faculty list serve to distribute invitations for the pilot survey. The desired sample size for the pilot survey was 100 participants who were representative of the total UNM population of 2688 faculty (University of New Mexico, 2007). The pilot survey did not include the 1873 faculty at NMSU (New Mexico State University, 2009).

Following the first round of individual and focus group interviews and the pilot survey, the next steps performed in parallel were: 1) developing descriptive statistics of the pilot survey data and 2) applying grounded theory to refine the constructs, dimensionality, and items to include in the final survey. Descriptive statistics consisted

of the means and standard deviations of the items included in the pilot survey along with the results of reliability analysis. Feedback from the interviews and pilot survey were coded and consolidated in the form of emerging results. The emerging results were presented to members of the dissertation committee and to additional interviewees to substantiate the reliability and validity of the data prior to finalizing the survey instrument. Faculty members who participated in follow-up individual interviews reviewed the pilot survey and provided suggestions on improving the content and format of the final survey.

The next step was administering the final survey to faculty members at UNM and NMSU by StudentVoice™. Appendix E includes an example of the final survey. Prior to the execution of the final survey, key faculty groups at UNM and NMSU were contacted to solicit their participation in the survey. The presentation to these groups included a description of the study and the emerging results based on the interview, pilot survey, and analyses conducted at that point. The UNM and NMSU Offices of the President and Offices of the Provost provided approval to use all faculty list serves to send out invitations for the final survey. The final survey was a census of approximately 1900 faculty members at NMSU and 2700 faculty members at UNM (New Mexico State University, 2009; University of New Mexico, 2007a). Paper copies of the survey were available as a backup to the web-based survey.

After completing the survey, the next step was analyzing the data using descriptive and inferential statistics. The primary technique used for analysis of survey data was principal component analysis (PCA) with the intention of finalizing the constructs and the dimensions that apply to assessing president performance. A

secondary purpose of PCA was to obtain data to develop hypotheses that could be tested in follow-up research. Another technique used to analyze survey data was multivariate analysis of variance (MANOVA). The main purpose of MANOVA was to investigate whether there were group differences in the perceived importance of the various items included in the survey. The reasons for investigating group differences were to obtain evidence of the validity of the scores and to gather additional information for the development of hypotheses for future research.

The next steps taken in the AIDA process was 1) preparing the statistical results, 2) making a final assessment of reliability and validity, and 3) and integrating qualitative and quantitative data to report the results of the study. Additional evidence of reliability and validity was obtained by conducting follow-up interviews to discuss the outcome of the final survey. While Figure 4 shows reliability and validity data collection and assessment as relatively late steps in AIDA, the dashed lines in the figure indicate that previous steps contributed data for reliability and validity assessment throughout the process. The final step in the process was summarizing the study findings (see Chapter 4) and preparing the discussion (see Chapter 5) which integrates, compares, and contrasts the results of the study with existing theory and research (Rudestam & Newton, 2001).

Data Management and Analysis

As a part of the AIDA model proposed in this study, the data management and analysis process consisted of a six-step process as illustrated in Figure 5. The first step was to develop the structure for the data files. The next step was to collect data for entry into the data files. Figure 5 illustrates data verification and validation as discrete, sequential steps; however, these steps were performed throughout the ADIA process as

parallel and iterative steps. After verifying and validating the data, mutually supporting qualitative and quantitative analyses were performed in support of a mixed-methods research design.

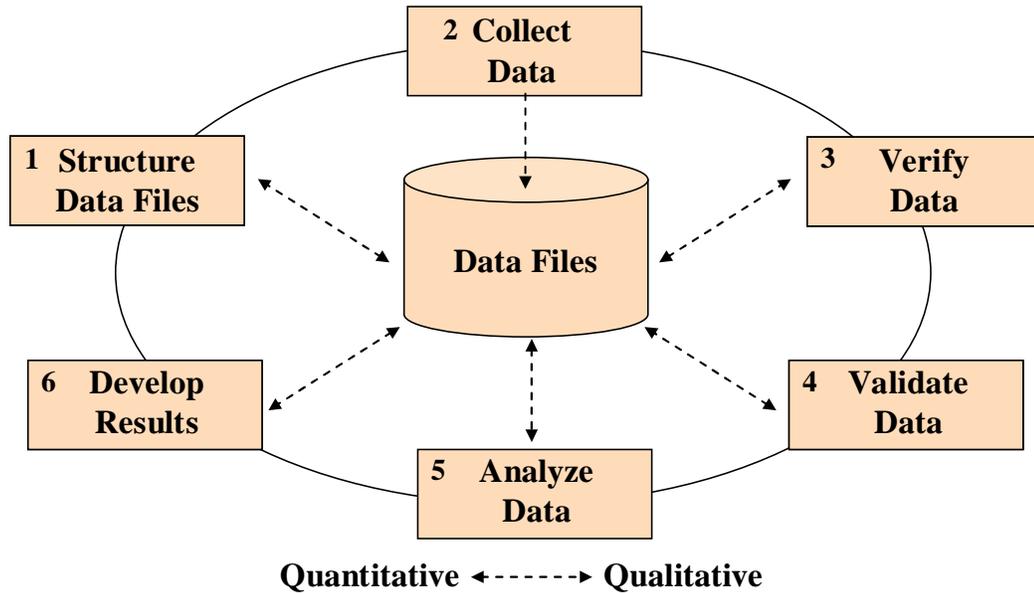


Figure 5. Data Management and Analysis Process

Data Files and Data Coding

The data files for this study consisted of Adobe[®] Acrobat[™] portable document format (.pdf) files, Microsoft[®] (MS) Excel[™] (.xls) files, MS Windows[®] Media Player[™] (.wmp) files, MS Word[™] (.doc) files, and Statistical Package for the Social Sciences (SPSS[®]) (data [.sav] and output viewer [.spo]) files. These files were maintained on a laptop computer and a flash memory drive and backed-up on compact disks (CDs). Qualitative and quantitative data were coded for retrieval and analysis purposes using the following schema:

- The first alphanumeric **A** stood for archival data, **I** for individual interview, **F** for focus group interview, **P** for pilot survey, and **S** for final survey.

- The second alphanumeric consisted of a set of numbers reflecting the data collection data (e.g., **04142009** for April 14, 2009)
- The third alphanumeric **a** stood for the first data file developed on a particular day, **b** for the second file, **c** for the third data file, and so on.
- For the final survey only, the following codes were used for survey questions to facilitate quantitative data analysis: **LN** for items corresponding to learning dimensions, **LE** for leadership items, **MN** for management items, **FO** for followership items, **OR** for organization items, **PA** for performance assessment items, **PO** for performance outcome items, and **EF** for external factor items.

Data Collection

Instruments

The qualitative data collection instruments for this study were archival data forms, individual interview guides, focus group interview guides, and a pilot survey questionnaire. These data collection instruments included questions pertaining to 1) presidential performance factors, 2) presidential assessment factors, 3) external factors that could affect presidential performance ratings, 4) and the content and format of the pilot and final survey instruments. The quantitative data collection instrument was the final survey questionnaire administered by StudentVoice™. The questions and format of this survey were modified based on the results of qualitative analysis of data from interviews and quantitative data from the pilot survey.

Sources

The primary sources of data for this study were libraries, web sites, higher education organizations, and faculty members at UNM and NMSU. Archival data forms (see Appendix G) were used to transcribe pertinent information from books, journal articles, newspapers, etc. In some cases, archival data forms were attached to relevant documents for organization and tracking purposes. The archival data form also served as a record for lessons learned in the application of the AIDA model for this study. Several universities provided information on their president assessment systems by responding to direct inquiries or by posting information on their public web sites. The Association of Governing Boards of Universities and Colleges (AGB) was a useful source of information about general policies, procedures, and studies associated with presidential assessment in colleges and universities. Due to the focus and scope of this study, faculty members were chosen as the target population for the individual interviews, focus group interviews, pilot survey, and final survey.

Purposes

The primary purposes of data collection included obtaining information on the constructs and dimensions of president performance assessment and on the relative importance of the variables derived from these constructs and dimensions. Another purpose was to obtain feedback on the pilot survey questionnaire and final survey questionnaire to improve the reliability and validity of the data collected from the administration of these instruments. The data collection process also served as a means to capture experiences and lessons learned from the application of the ADIA model in

this study. Finally, the pilot survey served as a source of data to rehearse the quantitative data collection and analysis procedures used in the study.

Table 2 summarizes the data collection instruments, sources, and purposes.

Table 2. Data Collection

Instruments	Sources	Purposes
Archival Data Form	<ul style="list-style-type: none"> ▪ Libraries ▪ Web sites ▪ Higher education organizations <ul style="list-style-type: none"> - Books - Journals - Newspapers - Board policy manuals - University procedures manuals - Past studies and interviews ▪ Researcher 	<ul style="list-style-type: none"> ▪ To capture data on applicable theories/concepts, president policies, processes, procedures, instruments, best practices, and issues at UNM, NMSU, and other comparable universities ▪ To capture data on the practical experiences and lessons learned in the application of the AIDA model
Individual Interview Guides	<ul style="list-style-type: none"> ▪ Faculty members 	<ul style="list-style-type: none"> ▪ To capture data on traits, behaviors, performance outcomes, practices, and external factors associated with university president assessment ▪ To obtain feedback on the content and format of the pilot survey and final survey questionnaire
Focus Group Interview Guides	<ul style="list-style-type: none"> ▪ Faculty members 	<ul style="list-style-type: none"> ▪ To capture data on traits, behaviors, performance outcomes, practices, and external factors associated with university president assessment ▪ To obtain feedback on the content and format of the pilot survey and final survey questionnaire
Pilot Survey Questionnaire	<ul style="list-style-type: none"> ▪ Faculty members 	<ul style="list-style-type: none"> ▪ To obtain qualitative and quantitative data analysis of factors associated with president assessment ▪ To obtain feedback on the content and format of the pilot survey instrument ▪ To collect data used to rehearse the quantitative data collection and analysis procedures for this study
Final Survey Questionnaire	<ul style="list-style-type: none"> ▪ Faculty members 	<ul style="list-style-type: none"> ▪ To obtain data for quantitative analysis of factors associated with presidential assessment ▪ To obtain feedback on the content and format of the final survey instrument

Procedural Details

The study included 15 individual interviews and 3 focus group interviews of faculty from UNM. Interview guides (Fraenkel & Wallen, 2006) were used to structure the questions and discussion during individual and focus group interviews. Appendix B contains the individual interview guides and Appendix C contains the focus group interview guides.

After obtaining consent from the interviewees, a written record was made of the responses to questions during the interview sessions and an audio recorder was used to capture responses. Once the interviews were completed, the written records were compared to the transcribed voice recordings to ensure data sets were accurate and complete. The written records and audio recordings were compiled on Grounded Theory Coding Worksheets (see Appendix G) to facilitate qualitative data coding and analysis.

Individual interviews were conducted during the pilot survey, after the pilot survey, and after the final survey. The objective of the individual interviews was to collect data on personal perspectives on president assessment factors and the quality of the survey instruments in an environment that offered an added degree of confidentiality. For individual interviews, invitations (see Appendix B) were sent to faculty members who had knowledge and experience in pertinent subject areas such as leadership, management, organizational development, assessment, and shared governance. Faculty members who did not respond initially were invited to participate in the study a second time.

Three focus group interviews were conducted – one early in the study and two near the end. The aim of the focus group interviews was to collect data on group and

individual perspectives in a more open environment than the individual interviews. For focus group interviews, invitations (see Appendix C) were sent to faculty members who had the same type of credentials as those invited to the individual interviews. For consistency, the questions and topics of discussion in the focus group interviews were the same as those in the individual interviews. The main reasons for conducting focus group interviews were to allow participants to develop or refute ideas introduced by others, to stimulate interactive discussion, and to reduce the influence of personal biases. The focus group interviews also provided data to compare and contrast with the data from individual interviews for reliability and validity assessment purposes.

The web-based pilot survey consisted of 118 total items of which 117 were closed-ended and one was open-ended (see Appendix D). Of the closed-ended items, 112 focused on the traits, behaviors, and outcomes associated with university president performance; the format of an assessment instrument; and external factors that could affect performance ratings. Five of the closed-ended items pertained to demographic factors (i.e., age, gender, race/ethnicity, category of faculty member, and employment outside the university). The open-ended item enabled survey participants to provide comments on president assessment and the pilot survey. The researcher conducted the pilot survey as a census (Fraenkel & Wallen, 2006) of the entire faculty (approximately 2700 total members) at UNM. The Offices of the President and Provost at UNM approved use of the all faculty list serve to distribute survey invitations via e-mail. Appendix D includes the invitation letter for the pilot survey. The use of StudentVoice™ with 106 participants was terminated after having achieved the goal of reaching at least 100 participants. The pilot survey participants provided 29 responses on additional

president assessment criteria and on ways to improve the final survey. The pilot survey included a summative attitude scale (Thorndike, 2005) in which the participants responded to statements by using a numerical indication of the importance of these statements in the survey. The numerical ratings ranged from “not important” (1) to “critically important” (5) across a graphical scale (Thorndike, 2005) placed to the right of the statements for ease of reference. For the pilot survey, $N = 106$.

Following the pilot survey, a *cross-sectional* survey was conducted of the total population of approximately 2700 UNM and 1900 NMSU faculty members. A cross-sectional survey is a survey administered on one occasion compared to a longitudinal survey administered on multiple occasions over an extended period of time (Fraenkel & Wallen, 2006). The sample frame (Dillman, 2007) was a list of individuals on the all faculty list serves at UNM and NMSU. Again, the Offices of the President and Provost at UNM and NMSU approved use of the list serves to send out survey invitations via e-mail. The e-mail included the uniform resource locator (URL) for the StudentVoice™ web-based survey. Appendix E contains an example of the final survey invitation letter.

The desired sample size for the final survey was a minimum of 420 participants from the combined UNM and NMSU population of 4600 faculty members. The desired sample size was determined by the number of questions in the survey to be analyzed using PCA with the desired number of participants being 5-10 per item (Field, 2005; Pett, Lackey, & Sullivan, 2003; Tinsley & Tinsley, 1987). The desired sample size for this study of 420 falls within the range suggested by Comrey (1973) of good (300 total participants) to very good (500 total participants) for factor analyses that produce interpretable and stable factors. Based on Monte Carlo studies, Guadagnoli and Velicer

(1988) found that stable factors were attainable using the following criteria: 1) at least three variables with $|.8|$ loadings on a factor regardless of sample size, 2) at least four variables with $|.6|$ loadings on a factor regardless of sample size, and 3) at least a few variables with $|.4|$ loadings on a factor with a sample size of 300 or greater. Given Guadagnoli and Velcier's guidelines, the desired sample size for this study should allow the researcher to use $|.3|$ as the suggested lower limit (Nunnally & Bernstein, 1994; Tinsley & Tinsley, 1987) for interpreting variables as part of a factor during PCA.

While the primary quantitative analysis method for this study was PCA, the researcher used MANOVA and multiple comparison procedures to determine if there were significant differences between the UNM and NMSU faculty for validity purposes. The pilot survey revealed effect sizes ranging from small ($\eta^2 = .01$) to large ($\eta^2 = .14$) (Cohen, 1988). Given the possibility of having small effect sizes for some items (variables) on the final survey, a sample size (n) = 386 would increase the likelihood of distinguishing the variables (up to 5 independent variables) with a significance level (α) of .05 and power ($1-\beta$) of .80 (Cohen, 1988). For the final survey, $N = 280$.

The final survey contained 64 items of which 42 supported PCA of learning, leadership, management, followership, organization, and performance outcome factors. The remaining items supported analysis of president assessment formats (7 items), external factors that can affect performance ratings (7 items), demographic factors (6 items), and the survey itself (1 item). With the exception of one open-ended item, the items in the survey were closed-ended. Similar to the pilot survey, the final survey incorporated a 5-point summative attitude scale (i.e., Likert scale) in a graphical format. Unlike the pilot survey in which the items were placed under categorical headings

(e.g., learning, leadership, management, etc.), the order of the items in PCA was determined by using a random number generator (Random.Org, 2009) to preclude any biases created by arranging items in hypothetical categories ahead of time. Appendix E contains the items and format of the final survey.

Immediately after completing the focus group interviews, individual interviews, pilot survey, and final survey, manual data collection forms and electronic media were stored in secure locations and input data files were transferred into a secure database for retrieval, processing, quality checking, and analysis. Throughout data analysis, subject matter experts (SMEs) provided feedback during interviews to confirm the trustworthiness of qualitative data as well as the reliability and validity of quantitative data. Checking the trustworthiness, reliability, and validity of data were continuous activities to identify and resolve issues as early in the study process as possible.

Consent forms were used for each data collection event involving human subjects. These forms identified the authority for the research, described the purpose and scope of the research, and provided information on the confidentiality of the data. These forms were provided to the participants for their own records and copies of the signed forms were kept in the research files separate from the data to prevent tracking of individual responses to names. Appendix F contains example interview and survey consent forms for this study approved by the UNM and NMSU Institutional Review Boards (IRBs). Appendix H includes the UNM and NMSU IRB exempt research approvals for this study.

StudentVoice™ (2008) was responsible for the development, administration, and summarization of the data for the web-based pilot survey and final survey. UNM researchers have access to StudentVoice™ assessment services free-of-charge because

the university is a member campus. The survey questions were sent to a StudentVoice™ technical representative who was responsible for formatting and automating the survey questionnaires. The surveys were available to the researcher for review prior to its administration to the faculty at UNM and NMSU. After finalizing the surveys, StudentVoice™ administered the survey. The researcher had access to actual data and summary data as soon as the first participant completed the survey. MS Excel™ and Adobe® Acrobat™ files of individual and summary data were available to the researcher during the survey and after StudentVoice™ deactivated the survey. Both the pilot survey and final survey were anonymous so there no way to connect the responses to individual names. The consent forms were provided on the first page of the survey and the participants had to acknowledge their consent before they could gain access to the survey questions. The survey participants had the option of answering some or all of the questions and discontinuing the survey at any time.

Data Verification

The data verification process for this study incorporated the following steps to ensure the data sets were accurate and complete:

1. Perform an inventory of all data sources and visually inspected databases to make sure that it contained data from these sources.
2. Ensure that data coding followed the schema specified for this study.
3. Back up all electronic data.
4. Identify missing data and rectify data losses.

- a. Perform Missing Value Analysis (MVA) in SPSS[®] to determine if missing values were missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) (Tabachnick & Fidell, 2007).
- b. Treat MCAR and MAR as ignorable non-response and pairwise deletion (which ignores the missing values of a particular subject for a statistical test) was used for statistical procedures (Alison, 2002).
- c. Treat MNAR as non-ignorable non-response and the imputation procedure of replacing missing cases with distribution means was applied since missing data did not exceed the often-used threshold of 15% of the data for a subject or a variable (George & Mallery, 2007). The replacement with the mean imputation procedure was also chosen because it is considered a conservative procedure since the variance of the variables is reduced (Tabachnick & Fidell, 2007).

Data Validation

It is essential to have a president assessment instrument that produces high quality data. According to Latham & Wexley (1994), the minimum standards for performance assessment instruments include incorporating a good job analysis and having good reliability and validity of data derived from the instrument. Job analysis identifies tasks that must be performed and standards that should be met. This study incorporated qualitative research methods in support of a job analysis for a university president. The associated tasks and standards provided a basis for survey items derived from the theoretical constructs and dimensions associated with university president assessment. Confirming the trustworthiness of qualitative data and the reliability and validity of

quantitative data helped ensure the survey instruments accurately measured the perceived importance of the traits, behaviors, and outcomes of a university president in the performance of assigned tasks.

Trustworthiness of Qualitative Data

For naturalistic inquiry, meeting certain tests of rigor is a requisite to establishing the trustworthiness of the outcome of an inquiry (Guba & Lincoln, 1981, 1985). There are four naturalistic (qualitative) corollaries to measures of quality in scientific (quantitative) inquiry. This research examined 1) *dependability* of qualitative data as an equivalent to the reliability of quantitative data, 2) *credibility* as an equivalent to the internal validity, 3) *transferability* as an equivalent to external validity, 4) and *confirmability* as an equivalent to objectivity.

An on-going audit was conducted during this study to obtain evidence of *dependability* by checking data consistency across subjects in the same or similar context. To ensure the *credibility* of qualitative data, several measures were applied during the study process as defined by Lincoln and Guba (1985) including:

1. Triangulation – comparing data from different sources such as observations, individual interviews, and focus group interviews to check for consistency and different viewpoints (i.e., data, methodological, and theoretical standpoints)
2. Peer debriefing – using disinterested parties to look at substantive, methodological, and ethical aspects of the study and provide feedback
3. Negative case analysis – resolving data that appears to be the exception rather than the rule by continuously revising hypotheses (or conclusions) until they account for all known cases without exception

4. Referential adequacy checks – holding a certain amount of data aside at various points during the research to determine if it is consistent with the findings, conclusions, and recommendations determined at the end of the research
5. Member checks – Formal, informal, and continuous checking of data and emerging results by subjects from whom data was collected over the course of the research

This study included *thick description* of the feedback from study participants, the research methodology used in the study, and the findings from the study as evidence of *transferability* of the results to other subjects or contexts. Denzin (cited in Creswell, 2007) describes thick description as a narrative that presents “detail, context, emotion, and webs of social relationships that evoke emotionality and self-feelings in which the voices, feelings, actions and meanings of the subjects are heard” (p. 194).

Evidence of *confirmability* was gathered by verifying the findings from the research stem from the characteristics of the research subjects and the context rather than from the biases, motivations, interests, and perspectives of the researcher. To support dependability, credibility, transferability, and confirmability, a reflexive journal that contained notes on perceptions, meanings, contexts, and potential biases was maintained during this study (Lincoln & Guba, 1985).

Reliability and Validity of Quantitative Data

As discussed in Chapter 2, reliability is the extent to which a measuring procedure yields the same results on repeated trials (Carmines & Zeller, 1979). Random error (unsystematic error) lies at the heart of reliability with higher random error reducing

reliability (Carmines & Zeller, 1979). In survey research, random error is a result of errors in coding, ambiguous instructions, and/or ambiguous wording on surveys. For this study, random error was minimized by checking code for errors as part of the data verification process and by conducting a pilot survey and follow-up interviews to clarify the instructions and the wording of the items in the final survey.

Cronbach's alpha (α) test of internal consistency was used as an indicator of reliability since it produces conservative estimates (Thorndike, 2005). Cronbach's α is based on the average correlation of items (Nunnally, 1978) within the survey instrument. The standard of reliability used for this study was Cronbach's $\alpha \geq .7$ as recommended by Nunnally (1978) for early stages of research focused on hypothesized measures of constructs. In addition, the number of non-demographic items was reduced from 111 on the pilot survey to 56 on the final survey. The results of the pilot survey and feedback from follow-up interviews enabled the researcher to identify those items that were more relevant for the final survey. Reducing the number of items on the survey also limited the artificial inflation of Cronbach's α , which is very sensitive to the number of items (N)

on an instrument based on the following formula: $\alpha = \frac{N^2 \overline{Cov}}{\sum s^2_{item} + \sum Cov_{item}}$ with \overline{Cov} as the

average covariance between all items, s^2_{item} as the sum of all item variances, and

$\sum Cov_{item}$ as the sum of all item covariances.

As described in Chapter 2, validity is the extent to which the scores of some abstract concept measure what they purport to measure (Carmines & Zeller, 1979). Nonrandom error (systematic error) lies at the heart of validity and increased nonrandom error causes the scores to be less accurate which lowers validity (Carmines & Zeller,

1979). The following types of validity were addressed in this study: 1) *statistical conclusion validity*, 2) *internal validity*, 3) *content validity*, 4) *construct validity*, 5) *external validity*, and 6) *consequential validity*. The researcher obtained the following evidence of validity to reduce the likelihood of nonrandom error scores captured from the final survey instrument and to increase confidence in the survey results:

1. Statistical conclusion validity evidence was gathered by calculating statistical power, conforming to assumptions of the statistical tests, controlling Type I error, and achieving reliability in the scores (Shadish, Cook, & Campbell, 2002).
2. Internal validity evidence was collected by conducting a cross-sectional survey to mitigate the following threats in survey research: mortality, location, instrument decay (Fraenkel & Wallen, 2006), and testing (Shadish, Cook, & Campbell, 2002). An additional threat to internal validity was the selection of participants with different characteristics, which could bias the scores (Shadish, Cook, & Campbell, 2002). Midway through the study, the NMSU faculty was added to the target population to compensate for any UNM faculty biases.
3. Content validity evidence was obtained by ensuring the content of the data collection instruments adequately incorporated the theories and concepts associated with the content domain of interest (Cronbach & Meehl, 1955; Lawshe, 1975; Shepard, 1993). Interviews with SMEs were performed to ensure the items pertained to the domain of university president performance and the items adequately sampled this domain. The literature review, follow-

up review of archival data, pilot survey, and interviews enabled the researcher to perform job analysis that identified the key tasks associated with university president performance. This job analysis contributed to the accuracy and inclusiveness of the items for the final survey serving as additional evidence of content validity.

3. Construct validity evidence was secured by analyzing threats such as inadequate operationalization of constructs (e.g., inaccurate or incomplete definition of constructs, dimensions, variables, and measures) (Shadish, Cook, & Campbell, 2002). Using factor analysis (Nunnally & Bernstein, 1994) ensured that the dimensions of a particular construct correlated highly with each other (*convergent validity*) and ensured the dimensions of different constructs did not correlate highly with each other (*discriminant validity*) (AERA, APA, NCME, 2004; Campbell & Fiske, 1959; Shadish, Cook, & Campbell, 2002; Shepard, 1993). In support of construct validity, the study accounted for the concept of *plausible rival hypotheses* (Shepard, 1993) or *construct confounding* (Shadish, Cook, & Campbell, 2002) by identifying several external factors that could affect the ratings of a university president that may be beyond his or her control such as attribution errors, organizational influences, political influences, and rater errors. While the constructs and dimensions of these external factors were not developed and analyzed, questions in the interviews and surveys were included to determine if these factors could be considered rival hypotheses or confounding constructs in the

actual assessment of a university president for follow-up research as opposed to explicit evidence of construct validity for this study.

5. External validity (Fraenkel & Wallen, 2006) evidence was obtained by conducting a census of the target population and by confirming the participation of representative faculty based on demographic data provided by UNM and NMSU. Using consistent settings and conditions for the collection of survey data and providing the same information to study participants contributed to external validity as well.
6. Consequential validity was supported by mitigating potential negative consequences of the proposed and actual use of scores (Messick, 1980: AERA, APA, and NCME, 2004). Ensuring the confidentiality of all study participants mitigated negative consequences. The pilot and final surveys were anonymous and the names of individual interview or focus group interview participants were not revealed.

Another type of validity is *criterion-related validity* (Thorndike, 2005). Criterion-related validity is based on using the scores from one instrument to predict the outcome (criterion) in another situation. There are also two broad classes of criterion-related validity: *concurrent validity* and *predictive validity* (Thorndike, 2005). Evidence of concurrent validity can be found by correlating the scores from one instrument to criteria measured at the same point in time. Evidence of predictive validity can be found correlating the scores from one instrument to criteria measured at a later time. Using president assessment as an example, the high correlation of scores from the annual president assessment and the actual performance of the university in meeting its goals for

that year is strong evidence of concurrent validity. If there is a high correlation of scores between an annual president assessment and the actual performance of the university for subsequent years, there would be strong evidence of predictive validity. Once a new assessment instrument has been implemented for a complete assessment cycle, the scores could be compared to the measures of university success to determine if the scores from the instrument provide good evidence of concurrent or predictive validity as cases of criterion-related validity. Based on the definition of criterion-related validity, its application is beyond the scope of this study. However, it is discussed in this methodology section because of its potential applicability in developing performance assessment instruments for university presidents.

Validity of Data in Mixed-Methods Research

Creswell and Plano-Clark (2007), Tashakkori and Teddlie (1998), and other authors suggest that mixed-methods research requires additional validity evidence because of the combination of qualitative and quantitative data. To minimize threats to validity of mixed-methods research data, the following actions were taken:

1. Qualitative and quantitative data were drawn from the same population (i.e., faculty members at UNM and NMSU).
2. For the most part, the qualitative and quantitative data collection instruments addressed the same questions.
3. A guide was developed to ensure equivalent translation of terms among qualitative analysis, quantitative analysis, theoretical/conceptual framework for the study, performance assessment terminology, and assessment instrument terminology (see Appendix A, Table A1).

4. Qualitative and quantitative results were presented side by side to illustrate where the two types of data were complementary and where they were different.
5. Individuals with different backgrounds were chosen to participate in the interviews, which is a preferred method for exploratory research.
6. A relatively large sample size was used for quantitative analysis compared to a relatively small sample size for qualitative analysis.
7. Major themes developed from CIT and grounded theory qualitative analyses were used as a basis for the final survey that provided data for quantitative analysis.
8. The survey instrument for quantitative data collection was refined through rigorous procedures including the analysis of pilot survey data and interview data.
9. The study addressed both qualitative and quantitative data validity.

Data Analysis

Qualitative Data Analysis

Qualitative analysis focuses on the beliefs, assumptions, situations, actions, and accounts (Charmaz, 2006) that play an important part in developing useful theories and models. For this study, qualitative analysis was useful in determining 1) *how* a presidential assessment instrument should be developed and *how* the assessment should be conducted, 2) *why* presidential assessment did or did not work well in the past, and 3) *what* should be done to make it work better in the future (Curnan, et al., 1998).

Qualitative analysis clarified and extended the results of quantitative analysis. For

example, qualitative analysis provided an indication of cause and effect relationships of factors that are perceived to contribute most to president performance outcomes as an extension of the survey results.

The qualitative analysis process incorporated continuous compilation and review of relevant archival information from studies, books, and articles that describe university president assessment and leadership themes, schema, or frameworks (Russ-Eft & Preskill, 2001). CIT procedures were used during individual interviews and focus group interviews. Qualitative data from the interviews and pilot survey served as sources of data for the application of grounded theory methods. Annotated data collection forms (i.e., archival data forms and grounded theory coding worksheets) were used to categorize and code data for qualitative analysis. Audio recorders were used to ensure written recordings were accurate and complete. The paragraphs that follow elaborate on CIT and grounded theory concepts and procedures.

CIT is “a set of procedures for collecting direct observations of human behavior in such a way to facilitate their potential usefulness in solving practical problems and developing broad psychological principles” (Flanagan, 1954, p. 327). The aim of CIT is to identify specific incidents of effective and ineffective behavior with respect to a particular activity. Sokol and Oresick (in Berk, 1987) suggest that CIT is useful in determining criteria for managerial performance assessment through what is called a “behavioral event interview” (p. 386).

For this study, CIT was used during individual and focus group interviews to identify behaviors that the subjects perceived are either effective or ineffective for a university president. For example, an item in the interview guide read, “Think of a time

when a university president has done something that you personally observed or heard or read about that in your opinion was an example of good behavior.” In contrast to the first statement, the interview guide contained a second related item substituting “poor” behavior for “good” behavior. The researcher used the responses to these types of items in the interview guides to refine the constructs and dimensions providing a basis for the items and the scales in the final survey.

Based on a qualitative research literature review, grounded theory is considered a useful approach for studying a phenomena or process (Creswell, 2007). Grounded theory emerged from the collaboration between sociologists Glaser and Strauss (1975) during the 1960s. In order to add structure to qualitative research, Glaser and Strauss developed a clear set of written guidelines for developing a theory that is “grounded” in data collected from the experiences, actions, interactions, and processes involving people. Later theorists such as Charmaz (2006) viewed grounded theory as a bridge between traditional positivist/postpositivist methods supported by quantitative analysis and interpretive methods supported by qualitative analysis. The bridge to which Charmaz refers is the methodology associated with pragmatism as defined by Creswell (2007). Figure 6 provides an illustration of the grounded theory approach used in this study as derived from writings of various authors (Charmaz, 2006; Charmaz as cited in Hesse-Biber & Leavy, 2006; Creswell, 2007; Glaser & Strauss, 1975; Strauss & Corbin, 1990).

Data for the initial coding phase was obtained through open sampling (individual interviews and focus group interviews) which was purposeful and systematic (Strauss & Corbin, 1990). As part of the conceptualization process, the researcher linked properties and dimensions to form the context that pertained to categories (or phenomena) (Strauss

& Corbin, 1990) under study such as learning, leadership, and management. Charmaz (2006) suggested the application of two coding schemes in this phase – line-by-line coding and focused coding. For this study, line-by-line coding helped avoid bias in data categorization. It also provided new ways to examine data and helped identify gaps. Focused coding was used to identify codes that continually appeared in line-by-line coding and to create broader categories for organizing data. Products from the initial coding phase, supported the next phase – axial coding. See Appendix A for definitions of coding terms.

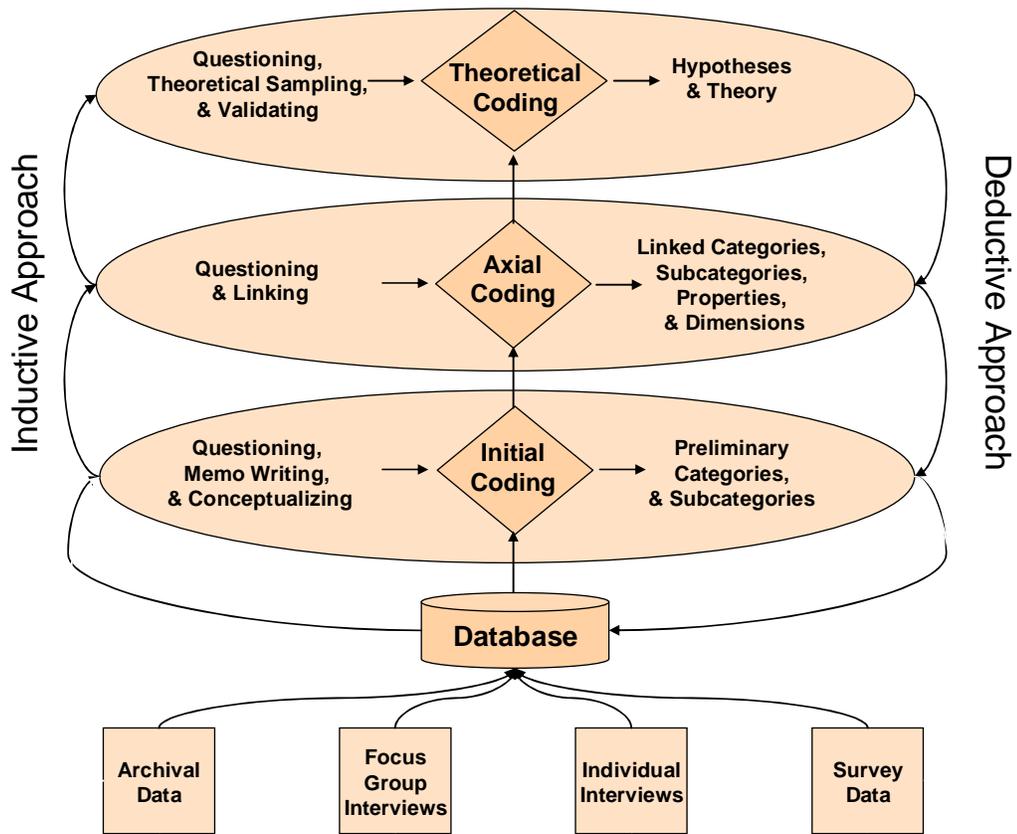


Figure 6. Grounded Theory Procedures

In addition to refining data by questioning its meaning, products were compared and linked from the initial phase during the axial coding phase. Detailed memos were

written during this phase in which subcategories were linked to preliminary categories. Products from the axial coding phase were candidate categories, subcategories, properties of the subcategories, and dimensions of the properties (Charmaz, 2006). The products from the axial coding phase were used during the next phase – theoretical coding.

The theoretical coding phase added theoretical sampling and validating as data processing procedures. Theoretical sampling involved going back to data sources to fill gaps and to differentiate major and minor categories providing an analytical framework for the hypotheses and theory derived from the grounded theory methods (Charmaz, 2006). Another key process performed in this phase included validating the data before writing the storyline, which identifies the core categories (or phenomena) of the study, the subcategories, and the relationships between these categories and subcategories. The preliminary products from the initial and axial coding phases were used to develop the final survey and the products of the theoretical coding phase were used to develop hypotheses on relationships that should be confirmed in follow-up research.

As Figure 6 suggests, grounded theory involves constant comparison of data throughout the research process. Data from one phase serve as building blocks for successive phases using bottom-up, inductive logic to develop more abstract concepts and emerging theory. Conversely, from a top-down, deductive perspective, the study included a review of theories and concepts identified during the literature review to determine if they were consistent with additional data introduced as the study progressed. As illustrated by the connecting lines in Figure 6, the grounded theory approach used in this study was an integrated and iterative process rather than a purely sequential process.

Quantitative Data Analysis

As a complement and extension of qualitative analysis, the researcher used descriptive and inferential statistics (Fraenkel & Wallen, 2006) for quantitative analysis. Statistical procedures were performed using pilot survey data to 1) exercise quantitative analysis procedures, 2) gain insight into the feasibility of using various quantitative procedures by examining the tenability of pertinent statistical assumptions, 3) identify items for the final survey and 4) assist in determining the sample size for the final survey. Pilot survey data was used to compute descriptive statistics (i.e., means and standard deviations) to prioritize potential items for the final survey and to determine the sample size necessary for MANOVA. While the pilot survey served multiple purposes, the data was not used to perform *PCA*, *ANOVA*, or *MANOVA* for quantitative analysis. Instead, data was used from the final survey to perform *PCA*, *ANOVA*, and *MANOVA* in support of this study.

After final survey data was entered into the database, it was screened to check for accuracy, completeness, and suitability. Raw data was examined to ensure data fields were complete and the values were within expected ranges. After confirming data quality, the data was input into SPSS[®] to obtain descriptive statistics and graphical plots to gain additional information on the quality and characteristics of the data. Outputs from SPSS[®] also included *histograms*, *box plots*, *distributions of means*, *Pearson's product-moment correlation (r) matrices*, and a variety of tables associated with *PCA*, *ANOVA*, *MANOVA*, *Sheffé's* multiple comparison procedure, and tests of statistical assumptions.

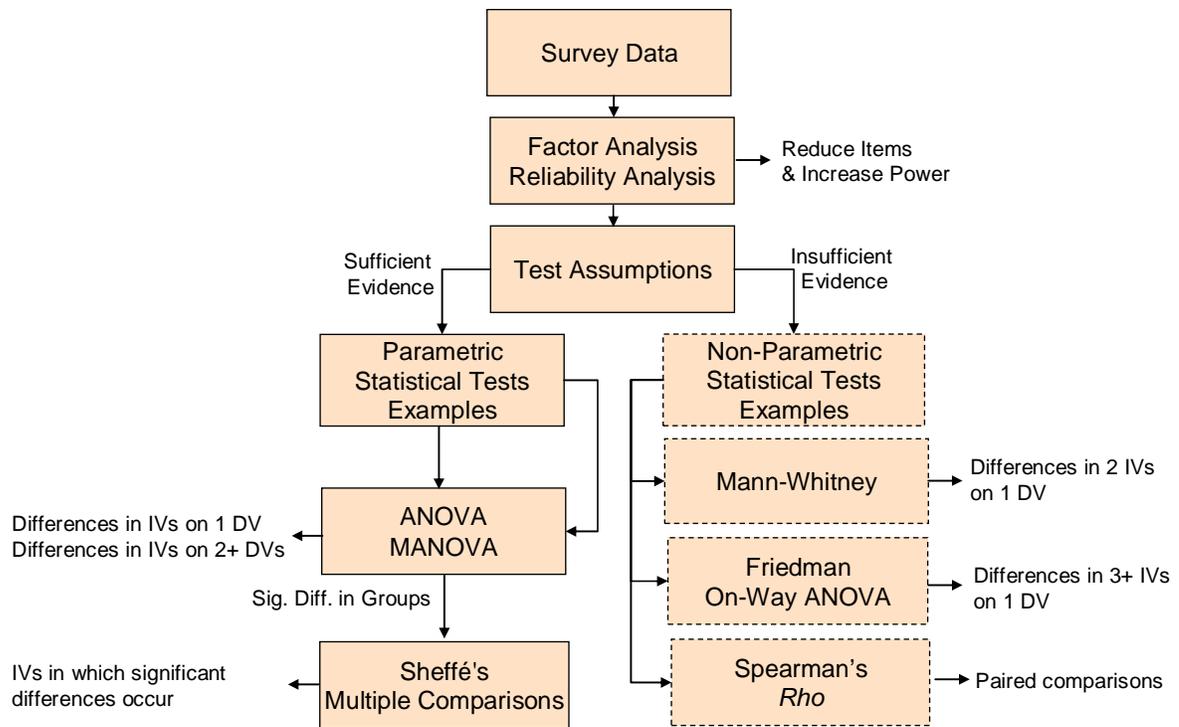


Figure 7. Statistical Testing Process

Figure 7 illustrates the statistical testing process used in this study. Factor analysis in the form of PCA and reliability testing in the form of Cronbach's α were the primary statistical procedures used for this study to gain insight into the constructs (*latent variables*) and dimensions (*manifest variables*) (Field, 2005; George & Mallery, 2007; Pett, Lackey, & Sullivan, 2003; Tabachnick & Fidell, 2007) reflected in the data from the final survey. The variables included in PCA were potential president assessment criteria associated with the constructs of learning, leadership, management, followership, organization, and performance outcomes. The items on the final survey corresponded to the variables associated with each of these constructs.

For PCA, tables in SPSS[®] output files such as the *total variance explained* provided insight into what components (or factors) should be retained and the percent of

variance accounted for by each component. Stevens (2002) suggests that the desire is to have the components that account for most of the variance (i.e., 75% or more) in the original set of variables, and this is often accomplished with five or less components. The *Kaiser criterion*, the most widely used component selection method, calls for retention of components with eigenvalues of 1.0 or greater (Weiss, 1971). Field (2002) suggests that the Kaiser criterion is accurate when the sample size exceeds 250 with the average *communality* of .6 or greater. Communality is the amount of variance a particular variable shares with other variables (Field, 2002). A communality of 1.0 indicates that a variable shares all its variance with other variables and a communality of 0.0 indicates a variable shares none of its variance with other variables (Field, 2002).

A *Scree Plot* (that graphs eigenvalues versus components) is another useful tool for identifying components to retain in PCA (Field, 2005). For the purpose of this analysis, the *break* in the Scree plot was examined and all components to the left break were considered as real components and those to the right were regarded as error or residual factors (Weiss, 1971). The break is considered a point in the plot where the slope of the curve tends to level off and Stevens (2002) recommends that all components be retained prior to the first point on the plot where the eigenvalues start to level off.

The *PCA components matrix* in SPSS[®] lists the unrotated loadings on the variables and the *rotated component matrix* lists the variable loadings from orthogonal (Varimax) rotations. Variable loadings less than $|.3|$ were suppressed and the following guidelines recommended by Guadagnoli and Velcier (1988) were used as other criteria for interpreting components:

1. Multiple loadings of $|\geq .8|$ regardless of sample size
2. Four or more loadings above $|\geq .6|$ regardless of sample size
3. Ten or more loadings above $|\geq .4|$ for samples size greater than 150
4. A few loadings of $|\geq .4|$ for sample size of 300 or more

Components defined by only one or two variables were considered questionable regardless of the loadings and sample size (Tabachnick & Fidell, 2007). Key matrices used to evaluate the outcome of oblique rotations (Promax) were the SPSS[®] *pattern matrix* (that consists of loadings of the variables on the components), *structure matrix* (that shows the relationships between factors), and *component correlation matrix* (that shows the dependence among component constructs). Both orthogonal and oblique rotations were performed to compare the results.

An important step in confirming statistical conclusion validity is testing the assumptions of statistical tests. According to Tabachnick and Fidell (2007), the key assumptions for PCA are as follows: 1) sufficient sample size for reliable correlation estimates, 2) multivariate normality (factor solutions are enhanced with multivariate normality, but this is not critical for exploratory research), 3) correlation matrix is not an identity matrix, and 4) absence of outliers (i.e., an observation is very different than most others [Field, 2002]). The *Kaiser-Meyer-Olkin* (KMO) test served as the measure of sampling adequacy with a goal of KMO value of at least .7 (Field, 2005; George & Mallery, 2007; Kaiser, 1974). *Bartlett's test of sphericity* was used to determine if the correlation matrix (*R*-matrix) was an identity matrix with a significance value $< .05$ indicating that the data is acceptable for factor analysis (Field, 2005; George & Mallery,

2007; Tabachnick & Fidell, 2007). Histograms and box plots were used to identify outliers (Tabachnick & Fidell, 2007).

MANOVA, ANOVA, and Sheffé's multiple comparison procedure were used to a limited degree for quantitative analysis in this study. The purpose of these tests was to determine if there were significant differences between groups in their perceptions of the importance of various performance assessment criteria. The independent variables (IVs) were groups represented by different demographic factors included in the final survey (e.g., age, gender, race/ethnicity, institution in which the faculty member is employed, etc.). The dependent variables (DVs) were various performance criteria associated with learning, leadership, management, followership, and organization, performance outcomes represented by corresponding items in the final survey. MANOVA, ANOVA, and Sheffé's were used to identify significant differences between UNM and NMSU faculty (IV) in their opinions on the importance of performance assessment criteria (DVs).

MANOVA was used to identify any significant differences in the groups based on the total set of DVs. Univariate ANOVA was used to identify individual DVs that contributed to distinguishing groups. Hotelling's T^2 and Sheffé's multiple comparisons were used to identify differences between individual groups on each of the DVs rather than the set of DVs. For MANOVA, the test statistic was *Wilk's lambda* (Λ) that indicates if the DVs significantly differentiate the groups. The ANOVA univariate F test provided an indication of whether the individual DVs differentiate the groups by themselves, but did not take into account correlation of the DVs to give a complete picture. For Scheffé multiple comparisons, $F_{\text{Scheffé}}$ is the criterion used as the test statistic (Keppel & Wickens, 2004).

The assumptions of MANOVA and ANOVA were tested in support of the assessment of statistical conclusion validity. The assumptions for these statistical tests are as follows (Field, 2005; Stevens, 2002): 1) MANOVA: independent observations, observations on the DVs follow a multivariate normal distribution in each group (i.e., multivariate normality), and the population covariance matrices for the DVs are equal (i.e., homogeneity of covariance matrices), and 2) ANOVA: independent observations, observations are distributed normally on the DV within each group, and population variances for the groups are equal (i.e., homogeneity of variance [HOV]). Since MANOVA and ANOVA are considered robust against Type I error (i.e., rejecting the null hypothesis when it is true [Aron & Aron, 2003]) when there are deviations to normality, and since SPSS[®] does not check for these assumptions (Stevens, 2002), other procedures to perform normality tests were not investigated. The Box's *M* test was performed to check the equality of covariance matrices, and *Levene's* test checked for HOV. For the assumptions of MANOVA and ANOVA, the Box test and *Levene's* test should be non-significant (i.e., $p > .05$) (Field, 2005; George & Mallery, 2007). For ANOVA, the inspection of histograms provided an indication of the normality of distributions, along with values of *kurtosis* and *skewness* between ± 1.0 , which considered excellent indicators of normality for most psychometric purposes (George and Mallery, 2007).

If the data did not adequately meet the assumptions for MANOVA and ANOVA, the *Mann-Whitney* and *Friedman One-Way ANOVA* were selected as alternative nonparametric tests to examine the differences between two groups, or three or more groups, respectively. Additional alternative tests included the *Wilcoxon Signed Rank test*

(a nonparametric equivalent of the paired samples *t* Test) and the Kruskal-Wallis *H* Test (a nonparametric equivalent of ANOVA for three or more data sets) (Wagner, 1992). Finally, the *Spearman's Rank Coefficient of Correlation* (r_s) was selected as a back up procedure to determine the strength of the relationships between two sets as paired comparisons (Wagner, 1992).

Mixed-Method Analysis

The mixed-methods approach used in this study involved CIT, grounded theory, and statistical procedures. The purpose of the final survey was to measure the perceived importance of various president performance criteria identified through the literature review, pilot survey, interviews, and synthesizing this data through the application of CIT and grounded theory. After performing statistical analysis of data from the final survey, quantitative findings and qualitative findings were compared and contrasted through triangulation. Additionally, the results of the qualitative analysis were used to amplify and clarify the results of the final survey. The desired outcome of the mixed-method procedures was to derive results that are more trustworthy from a qualitative analysis standpoint and to derive results that are more reliable and valid from a quantitative analysis standpoint.

Figure 8 provides an example of the way data from the qualitative analysis were mapped into the items to be included in the survey. In this example, the *category* defined by CIT and grounded theory corresponded to the constructs and the *subcategories* corresponded to the dimensions of those constructs as part of the theoretical and conceptual architecture of the domain of university president assessment.

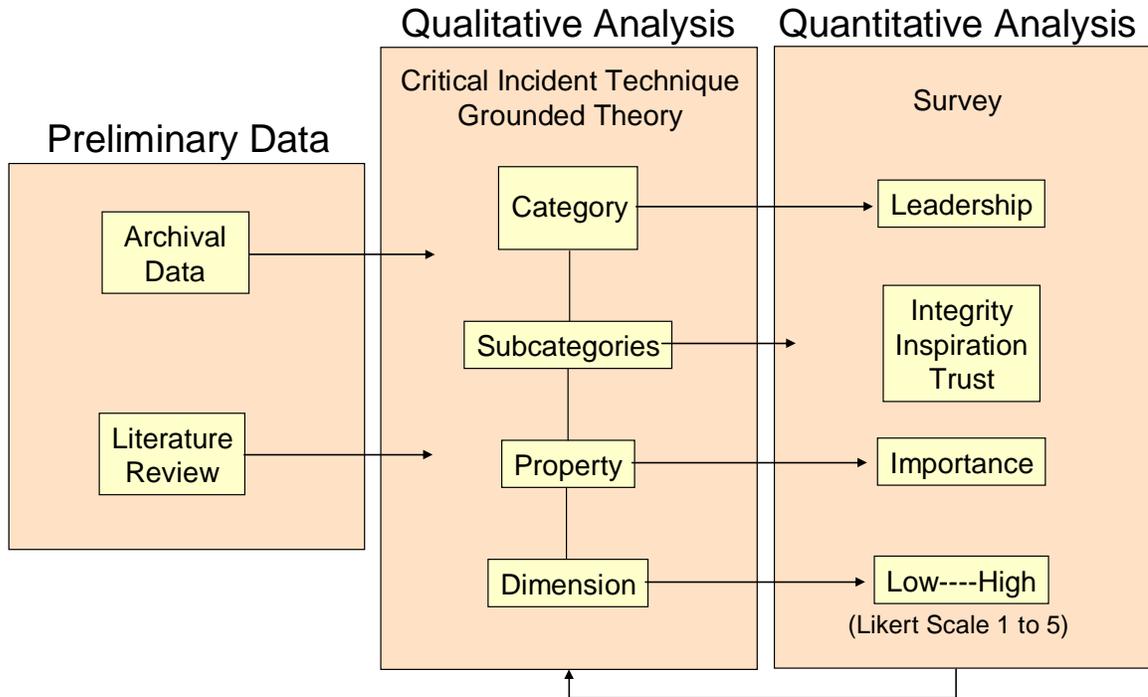


Figure 8. Example Linkage between Qualitative and Quantitative Analysis

The *property* in which this research focuses is the perceived importance of the subcategories (e.g., integrity, inspiration, trust, etc.) which are potential criteria for assessing the president. The term *dimension* from a qualitative perspective is equivalent to the range of a scale from a quantitative perspective. This approach of mapping from the qualitative perspective to a quantitative perspective was applied to determine categories, items, and scales that would be included in the survey instrument.

Risks and Mitigations

Several risks were taken into account during this study. Table 3 describes these risks and the actions taken to mitigate them.

Table 3. *Risks and Mitigations*

Risks	Mitigations
1. Lack of interest in participating as a study participant	1. Advertised and emphasized the benefits of the study.
2. Lack of interest in individual or focus group interviews	2. Advertised and emphasized the benefits of the study.
3. Insufficient participation in pilot surveys	3. Sent out invitations to participate via the UNM and NMSU all faculty list serves authorized by the Offices of the President and the Provost.
4. Insufficient sample size for the final survey	4. Engaged UNM and NMSU faculty leadership committees to solicit their support in recruiting survey participants.
5. Limited knowledge of senior leadership evaluation among stakeholders	5. Included multiple data collection methods in the study (i.e., individual interviews, focus group interviews, and surveys) to increase the chances of obtaining sufficient data. Interviews involved purposive sampling to improve chances that participants had knowledge and experience in applicable study domains.
6. Selection of an inappropriate statistical procedure for quantitative analysis	6. Conducted a pilot survey as a risk reduction measure for the final survey. Considered previous approaches used in related studies. Consulted with members of the dissertation committee to obtain additional insights into appropriate use of statistical tests. Collected evidence to support statistical conclusion validity.
7. Political influences that could bias the data and jeopardize research results	7. Included the faculty of two public universities in the state of New Mexico in the study. Compared and contrasted the results as part of validity assessment.

Summary

The methodology described in this chapter incorporates multiple aspects of exploratory research in a mixed-methods design. For this study, the main purpose of qualitative analysis was to explore constructs and dimensions of the university president assessment domain that provided a basis for the items to include in the final survey. The main purpose of quantitative analysis was to investigate relationships among the

variables represented by items in the final survey. Another purpose of the qualitative analysis was to clarify and amplify the relationships between the variables examined during quantitative analysis. While not a primary purpose, the mixed-methods design led to the development of hypotheses pertaining to the relationships between variables that could serve as a basis for follow-up research. Chapter 4 (Results) and Chapter 5 (Discussion) of this document address the findings, conclusions, and recommendations with regard to the AIDA model and the associated methods used to achieve the overarching goal of this study – to answer the following research questions:

1. What approach can UNM and other public universities use to develop an effective performance assessment instrument?
2. What is the preferred content and format for a president performance assessment instrument?

Although the target population was limited to UNM and NMSU faculty members for this study, the AIDA model and related methods could be applied to other populations to gain additional insight into the importance of various president assessment criteria. Feedback from other key constituents/stakeholder groups (e.g., members of the board of regents, administrative staff, and student body) could be incorporated into the design of a more effective assessment instrument. Once a president assessment instrument is developed, the AIDA model and methods described in this chapter could be applied to update the instrument on a periodic basis. Furthermore, the AIDA model and methods described in this chapter could be used to collect and to analyze data from an actual performance assessment of a university president.

CHAPTER FOUR

RESULTS

Chapter 4 contains the detailed findings from this study on the development of an instrument for assessing public university presidents. The theoretical and conceptual framework for this study provided a basis for the approach used to develop the results in this chapter and the discussion in Chapter 5. This chapter includes the following major sections: 1) problem statement and research questions, 2) restatement of the assessment instrument development approach (AIDA), 3) restatement of theoretical and conceptual framework, 4) findings on AIDA, and 5) findings on assessment instrument content and format.

Problem Statement and Research Questions

The findings in this chapter address the following problem statement and research questions:

Problem Statement

A comprehensive approach is not available that universities can use to develop the content and format of a president assessment instrument that provides sufficient evidence of the reliability and validity of the ratings or scores derived from this instrument.

Research Questions

- 1) What approach can UNM and other public universities use to develop an effective performance assessment instrument for their presidents?
- 2) What is the preferred content and format for a president performance assessment instrument?

Restatement of Assessment Instrument Development Approach

For this study, AIDA (Figure 9) served as a prototype approach for developing an effective performance assessment instrument for a university president to answer Research Question 1. Through the application of this prototype, the researcher captured lessons learned on the usability, utility, and limitations of AIDA as a model for identifying the content and format for an assessment instrument. The AIDA model embodies the methodological concepts described in Chapter 2 and Chapter 3 of this document. As such, it incorporates the methods and procedures associated with an exploratory, mixed-methods, triangulation, survey, and correlational research designs. As illustrated in Figure 9, AIDA includes a series of integrated data collection and analysis steps that lead to specific outputs.

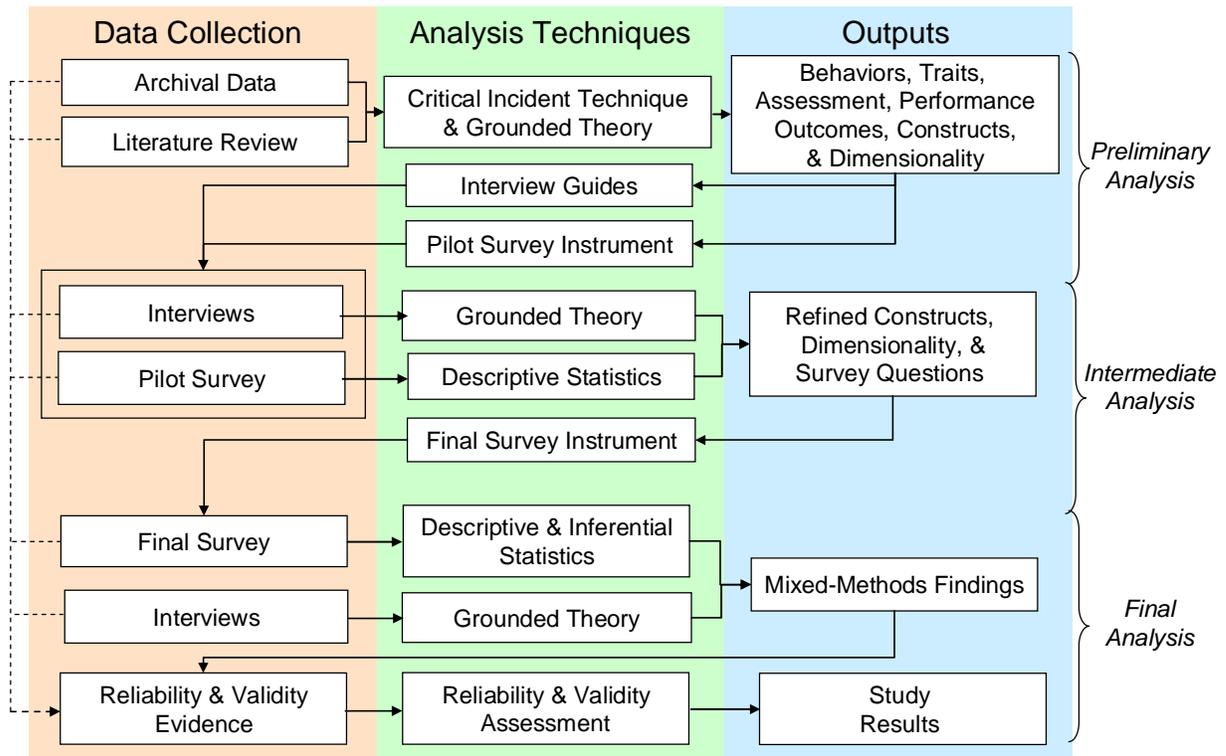


Figure 9. Assessment Instrument Development Approach

Due to the graphics limitations, AIDA appears as a linear process consisting of a series of sequential steps. In actuality, AIDA includes several interactive processes. For example, grounded theory entails a continuous comparison of initial, axial, and theoretical coding in a cyclical fashion to insure consistency and completeness of codes. Reliability and validity assessment is another example of a process that does not follow a linear process. The acquisition and assessment of reliability and validity evidence is an ongoing process since qualitative and quantitative data analyses are integrated activities in the AIDA model.

While AIDA consists of iterative and parallel steps, it also includes key sequential steps. For example, preliminary analysis for this study included qualitative methods (e.g., critical incident technique [CIT] and grounded theory) to determine the items to include in the interview guides. Preliminary qualitative analysis of archival data and information from the literature review also contributed to the development of the pilot survey. Following preliminary analysis, the researcher performed intermediate analysis of qualitative data from the interviews and quantitative data from the pilot survey. After intermediate analysis, the researcher conducted a final analysis of qualitative data from the follow-up interviews and quantitative and qualitative data from the final survey. Final analysis included integration of data from the final survey and follow-up interviews; assessment of reliability and validity; and reporting of study results.

Restatement of Preliminary Theoretical and Conceptual Framework

The preliminary theoretical and conceptual framework for this study (Figure 10) identifies factors associated with the assessment of a university president. The figure identifies existing theories that provide a basis for external factors, individual

performance factors, and organizational performance outcome factors. For example, assessment instruments currently in use by a cross-section of universities, literature, and studies reveal that specific items included in assessments of individual performance relate to learning, leadership, management, followership, organization, and assessment theories/concepts.

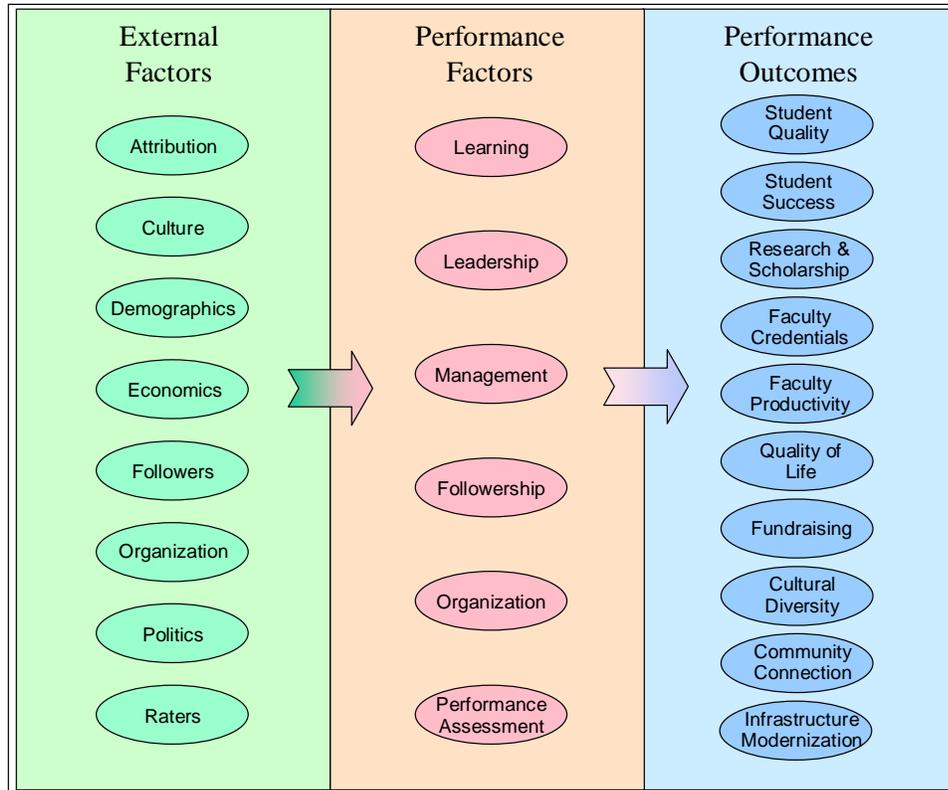


Figure 10. Preliminary Theoretical and Conceptual Framework

The preliminary framework for this study served as a point of departure for identifying the content and format for an assessment instrument that could potentially serve the needs of UNM, NMSU, and other public universities. Chapter 5 contains a refined theoretical and conceptual model and related hypotheses that should be tested in future research. Providing an answer to Research Question 2, Chapter 5 also contains the recommended

content and format of a UNM and NMSU president assessment instrument based on the perspective of key university constituents – the faculty members.

Findings on Assessment Instrument Development Approach

The following findings address Research Question 1: “What approach can UNM and other public universities use to develop an effective performance assessment instrument for their presidents?” The researcher developed AIDA because no model existed for developing the content and format for an assessment instrument for university presidents. While there is a significant amount of information on developing performance assessment instruments in general, much of this information is dated (e.g., Bernardin & Beatty, 1984; Mohrman, Resnick-West, & Lawler, 1989). Additionally, specific techniques for developing assessment instruments for senior leaders in public university environments are lacking. For example, previous research by Schwartz (1998) and others also point to the need for better definitions of university leadership and applicable measures of success.

Supporting Research Question 1, the following paragraphs in this section provide the findings from the application of the AIDA model in this study aimed at determining the appropriate content (i.e., president leadership definitions and measures) and format for implementing the instrument in a university environment. The findings that follow are essentially a critique of the AIDA model that the researcher applied during this study. Accordingly, these paragraphs address the usability, utility, and limitations of the various elements of the AIDA model. The findings associated with the outputs from the AIDA model are contained in the next major section of this chapter: Examination of Assessment Instrument Content and Format, which addresses Research Question 2. Chapter 5 of this

document includes recommendations on changes that should be made to the AIDA model based on the results of this study.

Findings from Preliminary Analysis

The main purpose of preliminary analysis was to identify the questions to include in the individual and focus group interview guides and the items to include in the pilot survey. Another purpose of this analysis was to begin gathering evidence of the dependability and validity of qualitative data from archival data sources and the literature review. Preliminary analysis also helped set the stage for more in-depth qualitative analysis performed in support of the intermediate analysis phase of this study.

Archival Data and Literature Review

Archival data on the demographics of the target populations at UNM and NMSU were readily available through the offices of institutional research or university websites. This data was useful in determining how closely the population of the pilot survey represented the UNM faculty population and the how closely the final survey represented the UNM and NMSU faculty populations. This demographic data was critical in determining the representativeness of the survey participants and the generalizability of the results as evidence of external validity.

Archival data in the form of president contracts, assessment instruments, policies, procedures, and performance reports for many institutions were easily accessible through the worldwide web. For example, the contract for the president of Arizona State University (ASU) (Arizona Board of Regents, 2005) contains specific language on performance assessment requirements and points to more detailed procedures in the Arizona Board of Regents (1990) policy manual. Both of these ASU documents were

available on the Internet. Other examples of university president contracts that contain information on evaluation are available to the public. Similarly, the University of Michigan (2006) and University of Washington (University of Washington Board of Regents, 2006) websites contain president evaluation forms.

Several universities and university systems post policies and procedures pertaining to president assessment on the Internet (Arizona Board of Regents, 1990; Clemson University, 2009; Krisch, 2008; McNeese State University, 2005; Southeastern Louisiana University, 1997; University of Alabama System, 2003; University of North Florida, 2006; Utah System Board of Regents; 2005). Some of these policies and procedures documents contain recommended president assessment criteria. The Utah System Board of Regents (2005) publishes comprehensive guidelines on president assessment that address evaluation committee procedures, evaluation processes, interview procedures, interview guides with detailed assessment criteria, and evaluation reporting.

There is also evidence that universities are taking actions to improve their policies, procedures, and assessment criteria. For example, the University of North Florida (UNF) chartered a president evaluation task force to develop “a plan that would provide for effective, regular evaluation of the University president” (University of North Florida, 2006, p. 1). The UNF task force delivered a concise report on the recommended policies, procedures, and criteria for presidential assessment. The task force based its recommendations on the review of relevant literature from the AGB and on the president assessment processes employed by 19 different colleges and universities including the Florida State University, State University of New York system, University of Florida, University of Massachusetts, and University of Minnesota.

A few universities make the results of president assessments available to the public. For example, the Board of Trustees of Ohio University (2006) published a performance report in the form of a letter to the president on the Internet. At one time, the University of Nevada Las Vegas (2000, 2003) posted detailed results of faculty and professional staff assessments of its president on a three-year basis. While states have “sunshine laws” (Kauffman, 1978, p. 64) that pertain to the release of public information such as the proceedings from boards of regents meetings, the lack of performance assessment reports on the Internet suggests that universities do not normally post results of president performance assessments on public websites. Additional information on New Mexico regarding the Open Meeting Act (OMA)/Inspection of Public Records Act (IPRA) (“sunshine laws”) is available on the New Mexico Attorney General (2008) website. Since states have their own laws and universities have different policies, there does not appear to be a common approach to release of information to the public including dissemination of performance assessment results. The Office of the President at UNM reports progress on goals and objectives approved by the board of regents on a quarterly basis (University of New Mexico, 2008). The UNM report is available on the Office of the President website and the Board of Regents uses this document for its annual assessment of the president (Salazar, 2009).

Based on the availability of information, it appears that there is a relatively high degree of transparency when it comes to universities releasing information on president assessment policies, procedures, and criteria. However, universities seem to be less willing to release the results of presidential assessments. Ingram and Weary (2000) point out that most states permit annual assessments to be confidential between the president

and board. They argue that “when ‘sunshine’ legislation or regulations adversely affect boards’ responsibilities to protect individual rights to privacy or to meet their responsibilities, public trustees and governing boards should aggressively advocate change” (p. 10). On the other hand, Ingram and Weary concede that president and board performance reviews conducted on more a general, longer-term basis (e.g., every five years) “should be more open to the public and inclusive, and the results should be made public in some appropriate form” (p. 10). Referring to their term “inclusive,” Ingram and Weary suggest that the longer-term assessment involves a special ad hoc board committee that seeks the “informed perceptions of leaders of major stakeholder groups – internal and external to the organization – concerning the organization’s management and governance” (pp. 11-12).

The literature review for this study revealed there was an extensive amount of literature on theories and concepts that can serve as a foundation for developing the content of an assessment instrument for senior leaders. On the other hand, there was a limited amount of recent and relevant literature on individual assessment at senior leadership levels – and even less literature on the subject of developing an assessment instrument for university presidents. Of the available literature on university president assessment, the AGB was the most useful source (see Ingram & Weary, 2000; Nason, 1997; Schwartz, 1998, 2001). Other beneficial sources of more general information on university presidents were the Center for Policy Analysis (2007), authors who are current and former presidents (see Bowen & Shapiro, 1998; Brown, 2006; Bruce, 2008; Duderstadt & Womack, 2003; Hoffman & Summers, 2000; Keohane, 2006; Rhodes, 2001), and authors who have credentials as subject matter experts (SMEs) (Alfred, 2006;

Padilla, 2005; Sanaghan, Goldstein, & Gavel, 2008). While these more general sources did not provide significant information on president assessment, they articulated job requirements, priorities, and performance expectations that could be applied in the development of an assessment instrument.

Consistent with Schwartz's (1998) suggestion that leadership definitions and corresponding measures are lacking for college and university president assessments, the researcher found no literature that provided an approach for developing a president assessment instrument based on applicable theories and concepts such as learning, leadership, and management. While some literature provided example criteria or measures for assessing university president performance, it did not include persuasive arguments for the inclusion or exclusion of various criteria or the format of the assessment. In fact, authoritative literature on the topic of university president assessment was contradictory when it came to the value of formal versus informal assessment and the value of private, one-on-one feedback versus participative, 360-degree feedback.

Despite any shortcomings of the review of archival data and literature, the information available for this study was very useful in development of the AIDA model. This information was critical in determining the content and format for the interview guides, pilot survey, and final survey for this study. The archival data and information from the literature review enabled the researcher to refine the theoretical and conceptual model providing a better understanding of the relationships between key variables in this study. Finally, this information was crucial in developing the recommended content and format for an assessment instrument that could serve the needs of a particular university.

Critical Incident Technique and Grounded Theory

CIT and grounded theory provided the basis for qualitative analysis for this study. Using archival data and information from the literature review, the researcher developed questions included in the individual and focus group interview guides and pilot survey by using the structured qualitative approaches specified in CIT and grounded theory. As a form of job analysis, the researcher used CIT procedures to formulate questions pertaining to behaviors, traits, and outcomes that contributed to the successful performance of a university president; and those that detracted from an individual's performance. As a parallel process, the researcher applied grounded theory techniques to identify *subcategories* of information through initial coding. These subcategories provided a basis for the interview questions and pilot survey items. Based on the coding of subcategories, the researcher used axial coding to identify properties and dimensions of the Likert scale applied to the questions in the pilot survey.

CIT and grounded theory proved to be valuable tools for developing questions and items for the interview guides and pilot surveys, respectively. CIT and grounded theory turned out to be relatively simple and straightforward techniques. However, developing a diagram such as Figure 6 in Chapter 3 helped the researcher understand the connection between the inputs, coding technique, and outputs.

Findings from Intermediate Analysis

The primary purpose of intermediate analysis was to refine the questions for the final survey. The secondary purpose of this analysis was to gather evidence of the dependability and validity of qualitative data collected during the individual interviews, focus group interviews, and the reliability and validity of quantitative data from the pilot

survey. Intermediate analysis set the stage for quantitative and qualitative analysis of final survey data and qualitative analysis of data from follow-up interviews performed in support of the final analysis phase of this study.

Interviews

The researcher conducted two rounds of interviews in support of the intermediate analysis phase. The first round of interviews consisted of ten individual interviews (with tenured and non-tenured faculty members) and one focus group interview (with five tenured faculty members). The main topics of discussion during the individual and focus group interviews were traits and behaviors of university presidents that contribute or detract from their successful performance. Additional topics included external factors that potentially affect performance ratings received by presidents, performance outcomes that may be good indicators of president success as well as the content and format of an assessment instrument. The researcher restricted the participation in the focus group interview to tenured faculty members as an added measure of protection due to the sensitivity of the subject. Another protective measure taken by the researcher was to hold in confidence the names and specific positions of the individuals who participated in the surveys.

As the primary source of qualitative data for this study, the interviews provided valuable information on the desired content and format of a president assessment instrument; key constituents/stakeholders who should be involved in the process; and factors that raters should take into account during an assessment that are beyond a president's control. The faculty members displayed a high level of interest in the topic and were very proactive in the discussion. Review of the individual and focus group

interview data revealed that faculty members had given 368 different responses to the questions included in the interview guides.

The researcher took notes and made audio recordings during the interview sessions. Immediately following the interview sessions, the researcher made written transcripts of the audio recordings on the grounded theory coding worksheets. Rather than using Atlas *TI*[®] to code interview data as described in the dissertation proposal, the researcher consolidated the data from the grounded theory coding worksheets to reduce the amount of time and effort necessary to analyze the data without sacrificing accuracy and completeness. The researcher found that the faculty provided clear and concise answers to the interview guide questions making initial coding and axial coding relatively simple and straightforward tasks for the intermediate analysis phase.

One of the major challenges was coordinating faculty participation in interviews. Multiple e-mails were required to obtain a response to the interview invitations. On the other hand, once the researcher established contact, the faculty proved to be very responsive. It took one month to conduct all the interviews with the exception of one, which took two additional months to make arrangements after several attempts to establish contact with different faculty members.

Pilot Survey

Even though the pilot survey consisted of items that could support more detailed quantitative analysis, the researcher limited its use to: 1) computing the means for each item and comparing them to interview findings, 2) gaining insight into the variability of data and the desired sample size for the final survey, 3) investigating the tenability of the assumptions for factor analysis (i.e., PCA) and MANOVA, and 4) rehearsing quantitative

data collection and analysis procedures. There were 106 participants in the pilot survey, 29 of whom provided comments on the indicators of university president success, the president assessment process, and/or the pilot survey. The researcher considered these qualitative inputs in the development of the final survey.

The researcher processed the pilot survey data using Statistical Package for the Social Sciences[®] (SPSS[®]) to obtain descriptive statistics to compare pilot survey and interview data and to determine the desired sample size for conducting MANOVA based on observed effect sizes. Considering the number of questions in the final survey and the desired sample size for factor analysis, the researcher determined the sample size for factor analysis exceeded the requirement for MANOVA. Using the pilot survey data, the researcher also performed factor analysis and MANOVA to check the tenability of the associated assumptions and to rehearse quantitative data analysis procedures for the final survey. Even though the sample size was relatively low for pilot survey data, the results were useful in gaining preliminary insight into the composition of the factors. The outputs from factor analysis also allowed the researcher to investigate how well the factors and items aligned with theory and concepts identified in the literature review. One of the problems discovered during this initial factor analysis process was the difficulty of interpreting the massive output from SPSS[®] when a large number of variables are included in the analysis.

The most significant challenge with the pilot survey was obtaining a timely response to the survey invitation. In order to achieve the goal of at least 100 participants, the UNM provost's office sent out two invitations via the all faculty list serve. Despite

the widespread dissemination of the invitation, it took approximately two months to obtain the desired number of responses to the pilot survey.

Data Analysis

During this analysis phase, the researcher consolidated the qualitative findings from the interviews and pilot survey, developed an MS PowerPoint™ presentation of these findings, and conducted a second round of follow-up individual interviews with three faculty members. The presentation of the emerging results emphasized similarities and differences in the findings from the interviews and pilot surveys. After presenting emerging results of the interviews and pilot survey, the researcher asked the interviewees to provide their opinions on the credibility of the results for validity purposes. The interviewees also reviewed each item on the pilot survey and made recommendations on what items to add, modify, or delete in the final survey. Using the results of the interviews and pilot survey, the researcher clarified the instruction for the survey, modified the wording of several questions, reduced the number of assessment-related questions from 111 to 56, and added one demographic-related question to the final survey instrument.

The emerging results presentation during this phase of analysis served two additional purposes. The first was to inform dissertation committee members of initial study findings and to obtain feedback on the direction of the study. The second was to inform members of key faculty committees at UNM and NMSU of the emerging results and to encourage participation of their faculty members in the final survey. The feedback from the interviewees, committee members, and faculty committee members suggested that the activities associated with this phase of analysis served their intended purposes.

Findings from Final Analysis

Final survey

The final survey proved to be a relatively effective instrument for collecting useful qualitative and quantitative data from UNM; however, the final survey was less effective at NMSU due to extended delays in data collection and low response. Since there was a large number (90 total) of responses to the additional comments question in the final survey, many of which were relatively long, Atlas *TI*[®] was used for content analysis (Klenke, 2008) as a supplement to the coding derived from grounded theory. Atlas *TI*[®] served as a useful tool for translating words, phrases, and sentences into a list of codes (Klenke, 2008). The list of codes derived from the final survey comments, along with the frequency count of the codes, provided insight into the central issues and themes expressed by survey respondents. Appendix K contains the codes for the final survey comments and the corresponding number of responses. Twelve final survey respondents provided feedback on the survey instrument. Some respondents commented on the vagueness of some of the questions and offered suggestions for improving the survey instrument.

Final interviews

The purpose of the final interviews was to obtain feedback on the results derived from the final analysis phase of the study for validity purposes. The researcher conducted two individual interviews and one focus group interview with three faculty members from UNM. Due to excessive coordination delays, the researcher decided to forgo final interviews with NMSU faculty members. The presentation to the final interview participants highlighted the integrated results and conclusions of the study.

The interview participants provided valuable comments on the credibility of the results and conclusions and offered suggestions on the interpretation of the major constructs identified through factor analysis.

Integration of Qualitative and Quantitative Data Analyses

The AIDA model was an effective means for integrating qualitative data and quantitative data at key points and for collecting evidence of reliability and validity throughout the research process. Having distinct, but integrated analysis phases allowed the researcher to condense large amounts of data into a more understandable format for the follow-up the presentations, interviews, and the final survey. The analysis phases enabled critical review and reflection at three logical points to facilitate resolution of themes by viewing results from both bottom-up (inductive) and top-down (deductive) points of view. Table 8 (Integrated Findings on Traits, Behaviors, and Performance Outcomes) provides an example of how qualitative and quantitative data was integrated to identify areas in which the two types of data were complementary or contradictory. Tables with integrated findings also provide supplementary information in the form data that was collected during the interviews or surveys (but not both) due to time constraints.

Assessment of Reliability and Validity

The AIDA model included reliability and validity assessment as an ongoing and iterative process in this study. Since this study focused on developing the content and format of an assessment instrument, reliability and validity were very important aspects of the research methodology. The reliability and validity process started with the review of a wide variety of authoritative literature on theories and concepts related to senior executive leadership, university president leadership, and performance assessment.

Faculty members with a broad range of experience and with varying degrees of contact with university presidents, boards of regents, senior administrators, constituents, and other stakeholders participated in the study to increase validity of the findings and results of this study. The qualitative and quantitative techniques outlined in Chapter 3 enabled the researcher to collect substantial reliability and validity evidence and to identify areas of improvement for future research on assessment of university president performance.

Presentation of Study Results

The researcher presented emerging results after completion of the preliminary, intermediate, and final analysis phases of this study. The presentation for the dissertation proposal articulated the preliminary results from the review of literature and archival data. The presentation on intermediate results served two valuable purposes – to validate the results of the initial interviews and pilot survey and to solicit participation in the final survey. In the case of UNM, the intermediate results presentation successfully served both of these purposes. Since the decision to add NMSU faculty in the study was made after the follow-up and their participation in the final survey was minimal, the intermediate results presentation did not fulfill either of these purposes at NMSU. The purpose of the final results presentation was to validate the overall study results including the final survey. Interviewees who participated in the final interviews were very receptive to the final results presentation and provided meaningful feedback for the dissertation manuscript that serves as the final report for this study.

The presentations of emerging study results proved to be a very useful element of the AIDA model. First, the presentations served as an efficient mechanism for collecting qualitative data that could be incorporated in the development of follow-up data

collection instruments. Second, they proved to be an effective advertising tool for the final survey at UNM. Finally, the emerging results presentations served as an effective means for obtaining additional evidence of reliability and validity for this study.

Summary of Findings on Assessment Instrument Development Approach

The AIDA model incorporated qualitative and quantitative analysis methods. As such, it required additional time and effort to apply the model in this study. However, Illgen and Favero (1985), Klenke (2008), and other authors argue that exclusive use of a quantitative paradigm for complex subjects such as leadership falls short because of the complexity of multivariate methods and the existence of complicated interaction effects makes it difficult to understand and interpret study results. Klenke insists, “quantitatively generated leadership descriptors often fail to lead to an understanding of the deeper structures of the phenomena we study” (p.4). Mason (2006), Creswell and Plano-Clark (2007), and many other authors cite the advantages of mixed-methods approaches such as increasing the capacity for theorizing, providing more diverse views, enhancing or extending the logic and explanation of relationships, and enabling researchers to make stronger inferences. For this study, the AIDA model facilitated looking at the data from multiple perspectives and provided a structured, building-block approach for integrating and reviewing findings throughout the study. One of its shortcomings was the process of conducting the pilot survey in parallel with the interviews. It would have been more useful to conduct the first round of interviews to incorporate the results into the pilot survey as suggested by Dillman (2007).

Findings on Assessment Instrument Content and Format

The following findings address Research Question 2: “What is the preferred content and format for a presidential performance assessment instrument?” In addition to applying the AIDA model to answer Research Question 1, it was also used for three complementary purposes to answer Research Question 2. The first purpose was to develop the content and format of the interview guides and surveys for qualitative and quantitative data collection in support of this study. The second purpose was to identify the content and format for a president assessment instrument based on faculty inputs from UNM and NMSU. The final purpose was to obtain data to refine the preliminary theoretical and conceptual framework for the study so the researcher could identify hypotheses pertaining to the variables within this framework for future research. The findings in the remainder of this chapter provide a basis for the discussion in Chapter 5.

Findings from Preliminary Analysis

The data used for preliminary analysis consisted of archival data and data collected during the literature review. Integrating the methods and procedures associated with CIT and grounded theory along with this data, the researcher developed the interview guides and the pilot survey. The paragraphs that follow highlight the findings from preliminary analysis that set the stage for intermediate analysis.

Archival data and literature review

The content of the pilot survey was derived from archival data and documentation studied during the literature review. The pilot survey contained demographic questions pertaining to faculty member gender, race/ethnicity, and employment status (e.g., full-time, part-time, assistant professor, associate professor, etc.) for comparison to data

provided by the UNM Office of Institutional Research to determine the representativeness of the pilot survey respondents. The demographic data on UNM faculty was readily available at the UNM website for the Office of the President. Additional demographic items were included in the pilot survey for future research that is beyond the scope of this study. Questions 9 through 14 in the pilot survey (see Appendix D) relate to demographics.

There was an abundance of information available to develop the 111 items included in the pilot survey derived from archival data and literature review documentation. The items in the pilot survey associated with the eight constructs of the preliminary theoretical and conceptual framework of this study (i.e., learning, leadership, management, followership, organization, performance assessment, external factors, and performance outcomes (see Figure 10), were also derived from archival data and documentation examined during the literature review. Items in the pilot survey were arranged under each of these constructs rather than being items being arranged in a random order. See the pilot survey in Appendix D for the specific items that relate to the eight constructs of the preliminary theoretical and conceptual framework.

The review of archival data and literature also provided a basis for the questions included in the individual and focus group interview guides (see Appendix B and Appendix C). The questions for the initial individual interviews and focus group interview addressed: 1) traits of a good president, 2) positive behaviors of a president, 3) negative behaviors of a president, 4) relevant president performance outcomes, 5) external factors beyond a president's control that can affect performance ratings, and 6) format for the assessment instrument in terms of formality, purpose, and recommended

participants in the assessment process. As opposed to the pilot survey that contained closed-ended items, the interview questions were open-ended, so the respondents were not restricted to the eight constructs identified through the review of archival data and literature. The responses to the open-ended interview questions served as a source of data to: 1) corroborate the findings from the pilot survey, 2) identify new items for the final survey, and 3) consider during follow-up interviews and analyses for this study.

Critical Incident Technique and Grounded Theory

CIT provided a structured technique for identifying the traits, behaviors, and performance outcomes of a successful university president from the perspectives of the faculty. For example, the initial individual interview and focus group interview guides included questions that required interviewees to reflect on instances that they had “personally observed, heard of, or read about” that were examples of good behavior (see Item 5 under Comments and Questions in the Individual Interview Guide in Appendix B). Consistent with the aim of CIT, these questions encouraged the interviewees to recall specific incidents of effective and ineffective behavior with respect to a president carrying out his or her responsibilities. The responses to the items in the interview guides were useful in refining the format and content of the final survey instrument for this study and in identifying candidate criteria to include in a performance assessment instrument for university presidents.

Grounded theory initial and axial coding served as a tool to identify the *categories* and *subcategories*. Strauss and Corbin (1990) define categories as an abstract grouping of concepts derived from qualitative data collected from research participants. For this study, *leadership* is an example of a category. A category coincides with a *construct* as

defined earlier in the context of the theoretical and conceptual framework. The individual concepts to which Strauss and Corbin (1990) refer are subcategories in grounded theory terminology. In the pilot survey, the individual items under each of the categories of learning, leadership, management practices, etc., are subcategories in grounded theory terms. For example, *integrity* is a subcategory under leadership in the pilot survey. The item in the pilot survey related to this subcategory is “Demonstrates a high degree of personal integrity” (see Appendix D, Item 2b).

In addition to defining categories and subcategories that provided a basis for pilot survey items, grounded theory served as tool to develop the attitude measurement scale (Oppenheim, 1966) for the pilot survey. Strauss and Corbin (1990) define a *property* as an attribute or characteristic of a phenomenon. Using personal integrity as an attribute or characteristic of president performance, its *importance* is a property associated with the phenomenon of leadership. Strauss and Corbin go on to define a *dimension* as a location of a property along a continuum. In terms of the property of importance, the dimension could range from high to low. See Appendix A for definitions of these grounded theory terms and Table A1 for a comparison of key terms used in this study.

Authors offer several options for anchors and Likert-type scales used in surveys (Dillman, 2007; Henning, 2007; Pett, Lackey, & Sullivan, 2003; Rea & Parker, 2005; Siegle, 2009). Applying grounded theory definitions for properties and dimensions for this study, the *anchors* (Pett, Lackey, & Sullivan, 2003) chosen for the survey measurement scale were “not important” and “critically important” with *ratings* (Oppenheim, 1966) of “1” and “5,” respectively. The *scale steps* (Pett, Lackey, & Sullivan, 2003) between these anchors were “very important” (“4”), important (“3”), and

“slightly important” (“2”). See Appendix D and Appendix E for the arrangement of the anchors and scale steps in the surveys.

Findings from Intermediate Analysis

Interview findings

The paragraphs that follow highlight the findings from the initial interviews and follow-up interviews. Appendix I contains detailed summaries of the responses to items during the initial individual and focus group interview that supported the intermediate analysis phase. Since the responses to follow-up interviews in support of the intermediate and final analysis phases were much narrower in scope, the findings are addressed in the following text rather than in a separate appendix.

Initial interview findings. Participants in the first round of initial individual interviews and the initial focus group interview responded to the same eight questions contained in the interview guides. Ten UNM faculty members participated in the first round of individual interviews and five faculty members participated in a focus group interview. The following paragraphs contain numbered items that indicate common responses from two or more interviewees. The more frequent responses are toward the beginning of the list and the less frequent responses are toward the end. See Appendix I for the entire list of questions and responses.

Question 1: What is the primary purpose of a presidential assessment?

Responses to Question 1: The interview participants provided 19 different responses to this question. The responses from two or more interviewees were as follows: 1) to provide inputs for improvement, 2) to assess success in the core mission,

3) to provide feedback on how well the president is doing, 4) to support retention decisions, 5) to set the tone for the university, 6) to provide accountability 7) to evaluate the performance of the president, and 8) to support personal development. The central themes for these responses appear to be related to personal improvement and personal accountability in accomplishing the core mission of the university.

Question 2: What are the traits of a good university president?

Responses to Question 2: The interviewees provided 70 different responses to this question. Some of the responses fell into the area of behaviors. Among the similar responses by two or more faculty members were: 1) honesty, 2) institutional knowledge, 3) effective communication, 4) empathy, 5) respect, 6) persuasiveness, 7) represents university well, 8) leadership, 9) integrity, 10) focus, 11) trust/trustworthiness, 12) optimism, 13) authenticity, 14) skill at interpersonal relationships, 15) courage, and 16) decision-making ability.

Question 3: What are positive behaviors of a university president?

Responses to Question 3: The participants provided 73 different responses to this question with several similarities and some overlap with responses to Question 2 pertaining to traits. The responses that interviewees had in common were: 1) shows appreciation and respect for others, 2) explains decisions, 3) serves as a mentor, 4) knows individuals/audiences, 5) has organization's best interests in mind, 6) argues the university's case to the State of New Mexico, 7) recruits and appoints the right people, 8) demonstrates courage and stands up to pressure/adversity, 9) shows willingness to sacrifice self-interests, 10) demonstrates accessibility/transparency, 11) makes cogent

arguments, 12) involves the faculty at the strategic level, 13) gets involved in the culture/State, and 14) demonstrates the ability to create a shared vision.

Question 4: What are negative behaviors of a university president?

Responses to Question 4: The interviewees gave 52 different responses to this question with many of the answers being the opposites of the positive behaviors or traits. The common responses were as follows: 1) demonstrates a lack of transparency, 2) acts dishonestly/deceitfully, 3) makes patronage appointments, 4) demonstrates self-centeredness/ties to own vision, 5) does not take inputs before making decisions, 6) acts defensively/takes things personally, 7) acts aloof/shows lack of connectedness, 8) tells people what to do, 9) lacks respect, 10) does not visit academic departments or get around campus, 11) adopts a corporate business model, 12) ignores advice, and 13) has too many goals upon which to act. The central themes in these responses appear to be related to lack of engagement and lack of due consideration of the opinions or perspectives of others.

Question 5: What performance outcomes should be included in an assessment?

Responses to Question 5: There were 44 different responses to this question. A high percentage of the responses related to higher-level goals and objectives of a university rather than those typically associated with president performance. The responses were as follows: 1) mood/climate/morale of faculty/organization, 2) resource acquisition, 2) university structure/systems development, 3) student-to-faculty ratio, 4) student success, 5) indicators that show movement in the right direction,

6) hiring practices, 7) resource allocation to the core mission (teaching and research), 8) UNM dashboard indicators with faculty/constituent inputs, 9) general and specific goals of the university, and 10) progress toward goals and objectives.

The faculty members brought up several considerations with regard to using institutional-level outcomes in president assessment. For example, more than one faculty member suggested that performance outcomes should be assessed as trends rather than a specific number at a given point in time. Multiple faculty members recommended that the performance outcomes should be focused on university strengths and should take into consideration factors such as the institutional climate, economy, and educational environment. More than one faculty member warned that it was important that there be transparency in the reporting of performance outcomes because the data could be manipulated to such an extent that they are inaccurate measures of the outcomes.

Question 6: What external factors affect the performance of a president that are beyond his or her control?

Responses to Question 6: There were 25 different responses to this question. The common responses by interviewees included the following: 1) economy, 2) regent's priorities, 3) decentralization and diffusion of authority with shared governance, 4) politics/political pressure 5) quality of incoming students, and 6) funding from the legislature. With regard to the quality of incoming students, more than one faculty member said that this is a controllable factor if the president invests sufficient time and effort in addressing this issue. Multiple interviewees also insisted that while presidents do not have direct control over external factors such as the economy, they are responsible for taking appropriate actions in response to these factors.

Question 7: Is formal assessment of value (with formal assessment defined as having written policies, procedures, assessment criteria, assessment instrument, and performed on a regular basis)?

Responses to Question 7: There were 49 different responses to this question. Compared to the others, this question prompted the most discussion during the interviews. With few exceptions, the interviewees were in favor of formal assessment for presidents; however, several of them posed the following suggestions for formal assessment: 1) it should be made public, 2) it should be anonymous, 3) it should be a 360-degree assessment, 4) it should involve multiple constituents, 5) it should have qualitative and quantitative aspects, and 6) its utility depends on its purpose and how it is accomplished. Two interviewees expressed skepticism of the value of formal assessment. One interviewee suggested that the institution should conduct an informal assessment at the beginning of a president's tenure and switch to a formal assessment if the informal approach is not serving a useful purpose. The same interviewee believed that informal assessment would be less expensive and be less of a distraction to the individual and the institution. Another interviewee was not convinced that an assessment of any kind would be of value at the level of a president or the university because there may too many different perspectives of what constitutes good and bad performance. The same interviewee expressed concern that an assessment could be a "club" as a form of punishment (Driscoll, 2000) that could jeopardize its usefulness. Despite the caveats and reservations, a high percentage of interviewees stressed that universities should perform president assessments on a regular basis – similar to faculty and staff assessments.

Question 8: Who should be involved in the assessment of a president?

Responses to Question 8: Interview participants provided 35 different responses to this question. The common responses by interviewees were as follows: 1) board of regents, 2) students, 3) administrators, 4) donors, 5) alumni/alumni board, 6) legislators, 7) deans, 8) faculty, 9) student leadership, 10) faculty senate, 11) staff, 12) vice presidents, 13) peer presidents, 14) president of the board of regents, 15) faculty senate president. While self-assessment appears to be an increasingly popular assessment practice, only one interviewee suggested self-assessment for a president. One interviewee stressed the importance of assessment feedback being anonymous, particularly for stakeholders such as staff members who may be concerned about retribution.

Follow-up interview findings. Three faculty members responded to five questions in the second round of individual interviews. The purpose of these follow-up interviews was to obtain feedback on the results of the first round of interviews and the pilot survey. The following paragraphs contain an abbreviated version of the question and responses by the interviewees.

Question 1: Have you taken the survey? If so, what was your impression of it?

Responses to Question 1: Two out of three interviewees had taken the pilot survey. Both of the interviewees who completed the pilot survey had favorable opinions of its content and format. One of the interviewees who had taken the survey said that it captured the critical aspects of president performance. The interviewee who had not taken the pilot survey was unaware that the invitation was sent out to the entire UNM faculty. This interviewee noted that it was easy to overlook e-mails sent out via list serves.

Question 2: Are the results of the interviews and pilot surveys reliable and valid?

Responses to Question 2: One interviewee commented that the results made sense, but disagreed with one of the findings. This interviewee was adamant that the president could have an impact on the quality of incoming students. Another interviewee believed the results made sense, but was surprised that faculty members rated certain items so high. This interviewee pointed out that his previous experience in shared governance might have influenced his perspective on the importance of various items in the survey. A third interviewee said that the results seemed “strong,” but the researcher should clarify the differences between leadership and management when presenting results in the future.

Question 3: How would you improve the format of the survey?

Responses to Question 3: None of the interviewees had significant issues with the survey format. One interviewee said the items were laid out in a logical format. Another interviewee said that the pilot survey was relatively long, which could reduce participation if the final survey were the same length. Another interviewee suggested that the researcher randomize the items rather than grouping them under major categories. To increase response rate, one interviewee recommended that the final survey be automated so the respondent can complete part of the survey, log out, and log back on to finish it at another time.

Question 4: Do you have suggestions on how to improve wording of the items?

Responses to Question 4: Each of the interviewees read the pilot survey questions and made suggestions on items to add, delete, or reword. One interviewee recommended that the final survey include an item that asks, “Who should administer the assessment?”

This interviewee also suggested that the researcher add a survey item to capture the number of years of experience as a faculty member at the university. One interviewee expressed concern that the survey may not contain all the pertinent items because participation was limited to faculty members. This interviewee stressed the importance of collecting data from other constituents and stakeholders such as the board of regents and alumni to ensure their perspectives are taken into account. One interviewee suggested that verbs be added to survey items associated with performance outcomes for the purpose of clarity. For example, “student-to-faculty ratio” should be changed to read, “reduces the student-to-faculty ratio.”

Pilot Survey Findings

Pilot survey quantitative findings. Over 100 UNM faculty members participated in the pilot survey. With the exception of the under representation of clinician educators, temporary faculty, and instructional faculty, participation in the pilot survey was moderately representative of UNM faculty demographics. Considering gender, race/ethnicity, and job categories (i.e., professor, associate professor, and assistant professor), the range in differences between the target population and pilot survey demographics was only one to six percent. Approximately 25% of the participants did not complete all the questions; however, the amount of missing data was less than 10%.

To prepare for quantitative analysis of data from the final survey, the researcher used the data from the pilot survey to rehearse applicable quantitative procedures using SPSS. The research reviewed descriptive statistics (e.g., means and standard deviations) for each of the items and performed exploratory factor analysis in the form of principal components analysis (PCA). Prior to performing PCA, the researcher tested for internal

consistency of the data and for the assumptions of PCA including 1) sample size adequacy, 2) multivariate normality, 3) absence of an identity correlation matrix (which is an indication of multicollinearity), and 4) absence of outliers. Cronbach's $\alpha = .96$ for the sample, which is very high. However, the very large number of items on the survey could have inflated this measure of internal consistency. For sample size adequacy, the Kaiser-Meyer-Olkin (KMO) test value was .78 indicating that the sample size was good (Kaiser, 1974).

Examining the factor loadings, many exceeded .512, which Stevens (1992) considered adequate for $n = 100$. In fact, many of the components had four or more loadings greater than .6, which Guadagnoli and Velicer (1998) consider to be reliable regardless of sample size. As an additional indication of the adequacy of the sample size, all of the communalities after extraction were greater than .6, which means that relatively small sample sizes (e.g., $n = 100$ or even less) may be adequate (MacCallum, Widaman, & Hong, 1999).

Bartlett's test of sphericity was significant ($p < .001$), indicating the correlation matrix was not an identity matrix (George and Mallory, 2007). The determinant of the correlation matrix was 9.82E-25. Considering that a value for the determinant $\leq 1.00E-5$ for a correlation matrix is an indication of multicollinearity, some researchers may consider eliminating highly correlated variables (e.g., $r > .9$) prior to proceeding with PCA (Field, 2005). However, Tabachnick and Fidell (2007) suggest that multicollinearity is not a problem for PCA because its computations do not involve the inversion of the determinant correlation matrix, which is problematic in the case of multicollinearity.

In addition to performing a test of internal consistency and PCA using pilot survey data, the researcher analyzed the data using MANOVA for practice. However, the low sample size and numerous small effect sizes (indicated by partial eta squared [η^2]), resulted in unacceptably low power for the multivariate tests. Furthermore, while the assumption of HOV was tenable for most DVs using Levine's tests, the tenability of the assumption of equality of covariance matrices could not be assessed using Box's *M* test because of the high degree of multicollinearity of the DVs. The reduction in the number of DVs for the final survey and the elimination of variables that do not correlate or that correlate very highly with other variables ($r > .9$) should reduce the problem of multicollinearity for quantitative analysis (Field, 2005).

Based on this preliminary investigation, there was sufficient evidence that the PCA procedures proposed for the quantitative analysis of final survey data were tenable for this study. However, a final determination on the utility of MANOVA was made using data from the final survey. Appendix J contains a table with the descriptive statistics of the scores for each item on the pilot survey.

Table 4 provides a list of the top twenty-five scores and the bottom twenty-five scores on president traits, behaviors, or performance outcomes with "5" being a "critically important" indicator of president success and "1" being a "not important" indicator of success.

Table 4. Highest and Lowest Scores for Traits, Behaviors, and Performance Outcomes

<i>Items</i>	<i>Mean Scores</i>
1. Displays a high degree of personal integrity (leadership)	4.75
2. Promotes institutional interests rather than self-interests (followership)	4.71
3. Builds trusting relationships with others (leadership)	4.63
4. Encourages open sharing of knowledge among constituents (learning)	4.58
5. Displays a high degree of job competence (leadership)	4.56
6. Adapts to changes that affect the university (learning)	4.35
7. Secures adequate resources (performance outcome)	4.34
8. Demonstrates professional courtesy to others (leadership)	4.33
9. Clears obstacles that enable constituents to be successful (leadership)	4.33
10. Student-to-faculty ratio (performance outcome)	4.29
11. Provides support to those individuals in leadership roles (followership)	4.29
12. Recruits high-caliber personnel (management)	4.29
13. Articulates university story (e.g., vision, mission, values) (learning)	4.24
14. Incorporates lessons learned into decision making (learning)	4.23
15. Establishes realistic goals for institution (organization)	4.18
16. Develops realistic plans to implement strategy (management)	4.18
17. Total revenue for the institution (performance outcome)	4.15
18. Funding from state appropriations (performance outcome)	4.03
19. Displays passion toward his or her work (leadership)	4.03
20. Uses appropriate performance indicators to make decisions (management)	4.00
21. Maintains good awareness of stakeholder satisfaction (organization)	3.99
22. Provides a framework for developing institutional strategy (leadership)	3.98
23. Serves as an agent for positive change (followership)	3.98
24. Rewards superior performance (learning)	3.97
25. Amount of research funding (performance outcome)	3.97
26. Emphasizes customer satisfaction (organization)	3.20
27. Amount of bequests received (performance outcome)	3.19
28. Percentage of minority students (performance outcome)	3.19
29. Student enrollment (performance outcome)	3.17
30. Number of faculty awards (performance outcome)	3.09
31. Number of degrees granted (performance outcome)	3.04
32. Focuses strategies on increased competitiveness (organization)	2.99
33. Student semester credit hours (performance outcome)	2.95
34. Number of student awards (performance outcome)	2.88
35. Tailors leadership style to follower expectations (leadership)	2.79
36. Number of Fulbright scholars (performance outcome)	2.69
37. Availability of extracurricular activities (performance outcome)	2.68
38. High school grade point average of incoming freshman (performance outcome)	2.65
39. Class standing of incoming students (performance outcome)	2.57
40. Number of transfers from other institutions (performance outcome)	2.56
41. Entrance examination scores of incoming students (performance outcome)	2.55
42. Rate of participation in extracurricular activities (performance outcome)	2.55
43. Number of international students (performance outcome)	2.35
44. Number of number of patents issued (performance outcome)	2.24
45. Acceptance rate of incoming students (performance outcome)	2.22
46. Number of license/option agreements (performance outcome)	2.21
47. Number of students studying abroad (performance outcome)	2.17
48. Number of start-up companies (performance outcome)	2.09
49. Success in intercollegiate athletics (performance outcome)	1.80
50. Total revenue from athletic events (performance outcome)	1.78

The words in parentheses in the items column in Table 4 correspond to the major categories of items in the preliminary theoretical/conceptual framework: performance

factors (learning, leadership, management, followership, and organization) and performance outcomes. Note that the items pertaining to performance outcomes do not contain verbs such as *reduces* the student-to-faculty ratio. Verbs were added to final survey items relating to performance outcomes in order to clarify the meaning of the item for reliability purposes.

Examining the top-twenty five items, there appeared to be a cross-section of items from the theoretical/conceptual framework that faculty members considered “very important” to “critically important” traits, behaviors, and outcomes that are indicators of president success. However, only three performance outcomes had mean scores that placed them in the top twenty-five. Similarly, looking at the bottom twenty-five items, twenty-two fell in the area of performance outcomes with faculty members classifying them in the general range of “slightly important” to “important” indicators of success.

External factors (i.e., attribution, culture, demographics, economics, followers, organization, politics, and raters) and performance assessment are additional major categories in the preliminary theoretical/conceptual framework; however, they do not fall under the areas of traits, behaviors, and outcomes as indicators of president success. Items falling under the category of external factors relate to influences outside the control of the president that could potentially affect his or her performance assessment ratings. Items falling under the category of performance assessment relate to the president assessment system itself (e.g., purpose, frequency, and formality). Accordingly, the findings for the items associated with external factors and performance assessment are included in separate tables that follow.

Table 5 contains a rank order of external factors that could potentially influence president performance ratings with a mean of “5” being a “critically important” factor and “2” being a “slightly important” factor. Considering the scores in Table 5, UNM faculty believed that factors such as shared governance, perspectives of the attributes of a good leader, conflicting goals/priorities, and scholarly criticism/skepticism were the more significant external factors. The researcher intended that Item 9 in Table 5 would serve as a statement that captured the relevance of the external factors in general. It is interesting to note how low the mean score of Item 9 is in comparison with the other items.

Table 5. Highest to Lowest Scores for External Factors

<i>Items</i>	<i>Mean Scores</i>
1. Shared governance with university stakeholders	4.15
2. Perspectives of followers on the attributes of a good leader	3.73
3. Multiple conflicting goals/priorities of the university	3.70
4. Scholarly criticism/skepticism within the university	3.67
5. Experience of stakeholders involved in the assessment	3.47
6. Economic conditions surrounding the university	3.37
7. Cultural backgrounds of university stakeholders	3.08
8. Rater errors (e.g., halo effect, leniency, and central tendency)	2.75
9. Variables in which the president has no control	2.66
10. Stakeholder desire to maintain autonomy from the staff	2.47

Table 6 contains a rank order of the preferred format of an instrument with a mean of “5” being a “critically important” format characteristic and a “3” being an “important” characteristic. Reviewing Table 6, UNM faculty placed relatively high importance on having multiple stakeholders involved in the application of the assessment instrument with the assessment being conducted on a regularly scheduled basis. They also rated the importance of having written policies, processes, and an assessment instrument that contains specific and objective criteria relatively high.

Table 6. Highest to Lowest Scores for Assessment Instrument Format

<i>Items</i>	<i>Mean Scores</i>
1. Involves multiple stakeholders in the assessment process	4.22
2. Includes assessment on a regularly scheduled basis	4.04
3. Documents associated policy/processes/procedures	3.80
4. Utilizes an assessment instrument with specific criteria	3.78
5. Provides ongoing feedback for personal development	3.76
6. Includes objective versus subjective assessment criteria	3.68
7. Ties president compensation to president performance	3.64

Table 7 is a summary of the ratings of the perceived importance of all items included in the pilot survey by the various categories (i.e., learning, leadership, management, followership, organization, assessment, external factors, and performance outcomes). The numbers in each row (e.g., learning, leadership, etc.) and column (e.g., “critically important,” “very important,” etc.) correspond to item numbers in the survey.

Table 7. Ratings of Pilot Survey Items by Category

<i>Category/Rating (mean score)</i>	<i>Critically Important (4.60 – 5.00)</i>	<i>Very Important (3.60 – 4.59)</i>	<i>Important (2.60 – 3.59)</i>	<i>Slightly Important (1.60 – 2.59)</i>
Learning	4	5, 9, 2, 3, 8	7, 15	
Leadership	11, 12, 13	20, 19, 21, 16, 18	10, 22, 14, 17	
Management		26, 23, 27, 24, 25, 28		
Followership	33	32, 34, 29, 30, 31		
Organization		36, 37, 41, 38, 40, 44, 35, 45, 42	39, 46, 43	
External Factors		62, 60, 61, 57	58, 59, 55, 63, 54	56
Performance		49, 50, 51, 52, 48, 53, 47		
Assessment		109, 92, 90, 91, 76,	69, 80, 75, 105, 111,	78, 86, 64, 87, 77,
Performance		103, 67, 102, 110,	101, 94, 112, 82, 95,	88, 106, 107
Outcomes		74, 68, 96, 108, 93, 104, 72, 97, 73	100, 84, 79, 70, 85, 71, 81, 98, 89, 65, 83, 66, 99	

The item numbers in each row are arranged in order from highest to lowest mean score.

Refer to Appendix J for a description of each item number contained in Table 7.

The arrangement of items in Table 7 suggests that pilot survey participants believed that items associated with learning, leadership, and followership are more important indicators of successful performance of a president than organizational

practices and performance outcomes. The pilot survey respondents considered that each of the format characteristics of an assessment instrument as “very important” to the successful assessment of the president and that several external factors may have a “very important” impact on president performance ratings. It is interesting to note that the respondents had lower ratings on performance outcomes, many of which are common performance indicators tracked by universities. The fact that there were a different number of items in the pilot survey under each category is relevant to the proper interpretation of Table 7. It would have been easier to interpret data in Table 7 if the number of items under each category had been equal.

Pilot survey qualitative findings. Twenty-nine survey participants provided feedback on the open-ended question that called for additional comments on indicators of university president success, the president assessment system, and/or the pilot survey itself. With respect to additional indicators of president success, respondents suggested that the president should: 1) listen and be responsive to constituents, 2) be committed to strategic planning, 3) care about the institution and its constituents, 4) support the core academic mission, 5) concentrate on student and faculty, 6) be respected by the faculty and students, 7) promote excellence in teaching and research, 8) not run the university using a corporate business model, 9) not cut services to students and departments to compensate for hiring new consultants, top administrators, and public relations people, 10) provide sufficient resources for basic school supplies (e.g., paper, pens, pencils, legal pads, and copies), 11) demonstrate innate leadership as the most significant hallmark of a successful university president, 12) possess cultural competency skills, and

13) demonstrate communication skills. Addressing the issue of president assessment, respondents provided the following replies: 1) an assessment instrument should allow for written comments, 2) an assessment of the context within which the evaluation is made should be done concurrently with president evaluation, 3) an assessment should include items pertaining to the president as an academic leader within and outside the university, 4) an assessment should include how he or she interacts with the board of regents, 5) the value of an assessment depends on how it is used and many people don't participate because they feel their voice will not be heard and it will not be worth the time, and 6) the assessment not be based on success in athletics. Finally, comments on the survey are as follows: 1) the survey does not consider current issues [at UNM], 2) it is not clear whether the survey is asking what the priorities are for assessing a president or is asking to assess the current president, 3) the survey should contain "no opinion" or "don't know options," 4) the first two-thirds of the survey was "wired" for the highest possible ratings, 5) some of the questions were "vague" and "obtuse," 6) the response scale did not fit all the questions, and 7) the survey should ask about the role of the board of regents.

Integration of Qualitative and Quantitative Findings from Intermediate Analysis

The aim of the individual interviews, focus group interviews, and pilot survey was to collect qualitative and quantitative data for addressing Research Question 2: What is the preferred content and format for a presidential performance assessment instrument? The paragraphs and tables that follow integrate the findings of the interviews and surveys and point out items on which the interviewees and pilot survey agreed were of relatively high importance. The integrated findings are addressed in the following order:

1. Preferred content of an assessment instrument based on interview feedback and pilot survey scores for traits, behaviors, and performance outcomes
2. External factors that can affect presidential performance ratings
3. Preferred format of a president assessment instrument based on feedback and pilot survey scores on the purpose, degree of formality, raters who should be involved, and frequency of the assessment

Table 8 provides an integrated list of traits, behaviors, and performance outcomes from the interviews and pilot survey. The words in parentheses in the items column in Table 8 correspond to the major categories of items in the preliminary theoretical and conceptual framework: performance factors (learning, leadership, management, followership, and organization) and performance outcomes.

Table 8 includes the mean scores for items that correlated between the interviews and pilot survey with associated scores. It also includes the exclusive responses in interviews and the exclusive responses in the pilot survey that participants rated on the average, as “very important.” Examining the findings on traits, behaviors, and performance outcomes, there is a correlation between the Items 1-16 with relatively high scores from the pilot survey and responses from the interviews. For Item 35, there was a strong correlation between the interviews and pilot survey on the importance of athletics and the need to avoid overemphasizing its importance. The interviews and pilot survey identified several additional traits, behaviors, and performance outcomes that may serve as useful content in a president assessment instrument. Items identified in the interviews and pilot survey comments were considered in the development of the final survey and the final theoretical and conceptual framework for this study.

Table 8. *Integrated Findings on Traits, Behaviors, and Performance Outcomes*

<i>Items</i>	<i>Interviews</i>	<i>Pilot Survey Mean</i>	<i>Pilot Survey Comments</i>
1. Demonstrates integrity (leadership)	✓	4.75	
2. Acts with organization's best interest in mind (followership)	✓	4.71	✓
3. Demonstrates transparency/openness/accessibility (learning)	✓	4.58	
4. Has strong job knowledge/competence (leadership)	✓	4.56	
5. Demonstrates ability to acquire necessary resources (management)	✓	4.34	✓
6. Displays respect/professional courtesy to others (followership)	✓	4.33	
7. Reduces student/faculty ratio (performance outcome)	✓	4.29	
8. Creates shared vision/tells university story (learning)	✓	4.24	
9. Establishes/makes progress on goals and objectives (organization)	✓	4.18	
10. Provides adequate resources (performance outcome)	✓	4.15	✓
11. Improves mood/climate/morale/satisfaction (organization)	✓	3.99	
12. Provides supporting structure/systems/framework for strategy (leadership)	✓	3.98	✓
13. Shows appreciation/rewards others (learning)	✓	3.97	
14. Increases student success/graduation rates (performance outcome)	✓	3.96	
15. Motivates/persuades/inspires others to follow his/her lead (leadership)	✓	3.76	
16. Demonstrates courage/raises controversial issues (followership)	✓	3.60	
17. Effectively communicates/understands audiences	✓		✓
18. Explains decisions/effectively argues case to others	✓		
19. Takes inputs from others for decision making	✓		
20. Gets involved in the culture/State of New Mexico	✓		✓
21. Connects with constituents/stakeholders	✓		✓
22. Does not become defensive/take things personally	✓		
23. Does not use a traditional corporate business model	✓		✓
24. Clears obstacles that enable constituents to be successful (leadership)		4.33	
25. Supports others in leadership roles (followership)		4.29	
26. Recruits high caliber personnel (management)		4.29	
27. Incorporates lessons learned into decision making (learning)		4.24	
28. Develops realistic plans to implement strategy (management)		4.18	
29. Displays passion toward work (leadership)		4.03	
30. Uses appropriate performance indicators to make decisions (management)		4.00	
31. Serves as agent for positive change (followership)		3.98	
32. Listens/responsive to constituents			✓
33. Concentrates on students and faculty			✓
34. Is respected by faculty and students			✓
35. Increases success in athletics (performance outcome)*	✓	1.80/1.78	✓

* The relatively low mean scores for Item 35 are indicators of the lack of importance of success in intercollegiate athletics and total revenue from athletic events.

Table 9 provides an integrated list of external factors that are beyond a president's control that can influence performance ratings along with the corresponding mean scores from the pilot survey. There were no comments from the pilot survey associated with external factors. Pilot survey mean scores ranged from "important" to "very important."

Table 9. *Integrated Findings on External Factors that Affect Performance Ratings*

<i>Items</i>	<i>Interviews</i>	<i>Pilot Survey Mean</i>
1. Decentralization/diffusion of authority with shared governance	✓	4.15
2. Perceptions of followers on attributes of a good leader	✓	3.73
3. Competing priorities (e.g., with the constituents)	✓	3.70
4. Scholarly criticism/skepticism	✓	3.67
5. Experience of constituents/stakeholders involved in the assessment	✓	3.47
6. Economic conditions	✓	3.37
7. Quality of incoming high school graduates (i.e., GPAs/class standings/exam scores)	✓	2.65/2.57/2.55
8. Politics/political pressure	✓	

Referring to Item 7, there appears to be some question as to whether the quality of incoming students is an uncontrollable factor and/or could have a negative impact on president performance ratings. Using grade point averages (GPA), class standings, and exam scores as measures of incoming student quality, the pilot survey scores indicate these factors may be more controllable by the president or have less of an impact on performance ratings compared to the other external factors. This finding is reinforced by interviewees who believed that a president has some control over the quality of incoming students. On the other hand, the lower scores on incoming student quality factors may indicate that pilot survey respondents generally believe that president performance ratings are not, or should not be, sensitive to these factors.

Table 10 provides a list of responses to interview questions that pertained to the format of a president assessment instrument along with the corresponding mean scores from the pilot survey and comments from the pilot survey. To reduce the length of the pilot survey, the researcher did not include questions pertaining to the desired participants in a president assessment. The researcher received recommendations on appropriate president assessment participants during the initial individual interviews and initial focus group interview.

Table 10. *Integrated Findings on Performance Assessment Instrument Format*

<i>Items</i>	<i>Interviews</i>	<i>Pilot Survey Mean</i>	<i>Pilot Survey Comments</i>
1. Assessment should involve multiple raters (e.g., 360-degree)	✓	4.22	
2. Assessment should be performed on a regularly scheduled basis	✓	4.04	
3. Assessment should include documented policy, processes, and procedures	✓	3.80	
4. Assessment instrument should have specific performance criteria	✓	3.78	
5. Assessment purpose is to provide inputs for performance improvement/personal development	✓	3.76	
6. Assessment instrument should include objective criteria	✓	3.68	
7. Assessment and president compensation should linked	✓	3.64	
8. Assessment should be context with the organizational climate	✓		✓
9. Assessment value depends on how it is used	✓		✓
10. Assessment purpose is to tell president how well he or she is doing	✓		
11. Assessment purpose is to increase accountability	✓		
12. Assessment purpose is to support retention decisions	✓		
13. Assessment purpose is to set tone for assessment in university	✓		
14. Assessment should be open and transparent	✓		
15. Assessment should be anonymous	✓		
16. Assessment requires balance of power among constituents	✓		
17. Assessment participants should include the president (self-assessment)	✓		
18. Assessment participants should include the board of regents	✓		
19. Assessment participants should include senior administrators (i.e., vice presidents and deans)	✓		
20. Assessment participants should include alumni/alumni board	✓		
21. Assessment participants should include faculty/faculty leadership	✓		
22. Assessment participants should include state officials (e.g., legislators, secretary of education, and governor)	✓		
23. Assessment participants should include students/student leadership	✓		
24. Assessment participants should include staff	✓		
25. Assessment participants should include donors	✓		
26. Assessment participants should include peers	✓		
27. Assessment should allow written comments			✓
28. Assessment should include items pertaining to academic leadership internal and external to the institution			✓

Summarizing the findings on the format of an assessment instrument, the interviewees and pilot survey respondents believed it was most important for multiple constituents and stakeholders to participate in president assessment. In terms of frequency, the interviewees and respondents indicated that it was “very important” for the assessment to take place on a regularly scheduled basis. The relatively high scores on requirement for documented policy, processes, procedures, and specific/objective criteria

suggest that the interviewees and pilot survey participants preferred formal assessment as opposed to informal assessment. With respect to purpose, the findings from the interviews and pilot survey indicated that the primary purpose of a performance assessment is to provide feedback to the president for performance improvement and personal development. In addition to providing specific recommendations on assessment participants, the interviewees made several other recommendations pertaining to the format for applying the president assessment instrument.

Summary of Findings from Preliminary and Intermediate Analysis

The first round of individual interviews and the focus group interview provided a significant amount of qualitative data on the preferred content and format of a president assessment instrument. Examining the data in Tables 8-10, there is considerable evidence of convergent findings based on the correlation of many pilot survey items with interview responses. There are also indications that the pilot survey and interviews were complementary and led to the development of a more complete picture of assessment instrument content and format because each research procedure identified additional items for consideration.

Referring to the integrated findings in Table 8, the faculty tended to identify positive traits and behaviors as more important indicators of president success compared to positive performance outcomes. The large number of performance outcome scores ranking in the bottom twenty-five of the pilot survey (see Table 4) reinforces this finding on the relative importance of trait and behaviors versus performance outcomes. Looking specifically at traits and behaviors, items related to learning, leadership, management, and followership items occurred more frequently on the list compared to items associated

with organization. With regard to performance outcomes, faculty indicated that increasing resources availability and student success are “very important” indicators of president success and potential assessment criteria.

While Table 8 suggests there are a wide variety of traits and behaviors to choose from when developing the content of a president assessment instrument, it also indicates that relatively few performance outcomes ranked high on the list of candidate assessment criteria. Findings from the interviews reinforced that faculty members had some reservations about using performance outcomes as assessment criteria. For example, multiple interviewees were not convinced that a president should be responsible for university-level performance outcomes at a “snapshot in time” because there were so many intervening variables.

Rather than making the president accountable for performance outcomes with specific numerical objectives by a certain time (e.g., reduces the student/faculty ratio by twenty-five percent in two years), some interviewees proposed that the president be accountable for showing progress toward achieving measurable objectives over time. Some interviewees said it was acceptable to use performance outcomes as assessment criteria, but stressed that those external factors, such as the institutional climate and economy, must be taken into account. A few interviewees warned that the development and assessment performance outcomes at a university level must be open and transparent to prevent distortion and misuse of the data.

Looking at integrated findings in Table 9, there was general agreement among the interviewees and pilot survey respondents that some external factors are beyond the control of the president and/or may have significant influence presidential performance

ratings. Specifically, the study participants identified the following external factors as potentially having significant impacts on ratings: 1) shared governance, 2) differing perceptions among constituents or stakeholders of the attributes of a good leader, 3) competing priorities among constituents or stakeholders, 4) scholarly criticism or skepticism attributable to the academic environment, 5) experience level of individuals involved in the assessment process, 6) economic conditions, and 7) political pressure. The lower mean scores on high school GPA, class standing, and entrance examination scores imply that president performance ratings are less sensitive to the quality of incoming students and/or may be a more controllable factor. Although the interviews and pilot survey results indicated that several external factors could affect performance ratings, multiple interviewees said a president is responsible for reacting to these factors in an appropriate manner. For example, multiple interviewees emphasized that while a president is not responsible for local, state, and national economic conditions, he or she is responsible for taking appropriate actions to mitigate any adverse effects.

Reviewing the integrated findings in Table 10, the faculty believe that it is “very important” that president assessment involve multiple raters and that it be conducted on a regular basis (e.g., annually). The relatively high scores on the following attributes of a performance assessment system imply that it is very important for the university to have a formal system rather than an informal system that: 1) requires assessment on a regularly scheduled basis, 2) includes documented policies, processes and procedures, 3) includes specific performance criteria, 4) includes assessment instrument with objective criteria, and 5) links assessment results and compensation decisions. Turning to the purpose of an assessment, the predominant theme is the importance of this process in providing

constructive feedback for personal development. Comments that assessments should be used by the board of regents to make compensation and retention decisions reinforce their use as an accountability tool. With respect to participants, the interviewees suggested that a wide variety of constituents and stakeholders be involved in president assessment.

Despite the relatively strong support for 360-degree, regularly scheduled, and formal assessment with a wide variety of participants, two of the interviewees had reservations about formal assessments. One of the interviewees suggested that a formal assessment instrument be used only if informal assessments were not having the desired effect of improving president performance. The same interviewee suggested that informal assessment would be less expensive and less of a distraction if they involved the “right people” (i.e., members of the board of regents, peer presidents, past presidents, vice presidents, deans, head of staff council, faculty committee leaders, and students organization leaders). Another interviewee was not convinced that any type of assessment would be of use to the president or the university unless there was an appropriate balance of power among those involved in shared governance (i.e., the board of regents, president, and faculty leadership). The same interviewee expressed concern that assessment can be an inappropriate form of punishment and demean a president’s position as chief executive officer of the institution.

Findings from Final Analysis

The following section addresses the quantitative and qualitative results from the final survey administered from September 2009 through January 2010 and final interviews conducted in March 2010. This section also merges the findings from the preliminary and intermediate analysis phases of this study.

Final Survey Quantitative Findings

The following paragraphs describe the quantitative methods and findings from analysis of final survey data. The final survey instrument contained 42 closed-ended questions in which participants rated the importance of various criteria for assessing university president performance. Item analysis and factor analysis (PCA) were used to explore the dimensions and constructs associated with these assessment criteria. Cronbach's alpha computed as a measure of internal consistency for the total 42-item scale as well as the factors derived from PCA. This section also describes the interrelationships of the factors to provide a broader perspective of the perceived importance of various president assessment criteria and to refine the original theoretical and conceptual model for this study as a basis for future research. Pett, Lackey, and Sullivan (2003) suggested that Leske's (1991) journal article on the Critical Care Family Needs Inventory provides "an exemplar of a published report on development of a [family needs] instrument" (p. 237). The following section incorporates Leske's suggestions for describing the methods and findings developed from item analysis and factor analysis.

This section also includes the mean scores for the: 1) forty-two items in the final survey pertaining to assessment instrument content, 2) seven items related to external factors that can influence president assessment ratings, and 3) seven items associated with president assessment format. Tables with means scores provide faculty perspectives on the relative importance of the items included in each of the three categories.

Sample. The final survey was completed by 189 faculty members from UNM and 91 faculty members from NMSU for a total $N = 280$. Table 11 contains the demographic

characteristics of the final survey participants. Based on demographic data, the highest percentage of participants were in the 50-59 years of age range, female, white, full-time professors with over 20 years teaching experience in higher education. Approximately two-thirds of the final survey participants were UNM faculty members.

Table 11. Demographic Profile of Survey Participants

<i>Variable</i>	<i>N</i>	<i>Percent</i>
Age (yrs)		
20-29	6	2.2
30-39	23	8.3
40-49	64	23.2
50-59	127	46.0
60+	56	20.3
Gender		
Male	126	46.0
Female	148	54.0
Race		
White	213	80.4
Black	2	0.8
Hispanic	24	9.0
Other	26	9.8
University		
UNM	184	66.9
NMSU (4-year)	80	29.1
NMSU (2-year)	11	4.0
Employment Category		
Full-time	153	88.4
Part-time	20	11.6
Assistant Professor	24	12.8
Associate Professor	44	23.4
Professor	69	36.7
Other	51	27.1
Years Teaching in Higher Education		
Less than 5	23	8.5
5-9	39	14.3
10-19	89	32.7
20+	121	44.5

Procedure. The Office of the Provost and UNM and NMSU authorized that an initial and follow-up invitation for the final survey be sent out via their all faculty list serves. *StudentVoice*TM administered the web-based final survey from September 2009 through January 2010. This study was approved as exempt research by the Institutional Review Boards (IRBs) at UNM and NMSU. The following Predictive Analytics

SoftWare[®] (PASW[®]) Statistics 18 (also known as SPSS[®]) subprograms were used for this analysis: 1) Descriptive Statistics (frequencies, descriptives, and cross-tabs), 2) Scale (reliability), 3) Dimension Reduction (factor), 4) General Linear Model (multivariate), and 5) Missing Value Analysis. These PASW[®] subprograms provided the following types of data: 1) participant demographics, 2) mean scores, 3) standard deviations, 4) principal components, and 5) validity data (e.g., tenability of statistical test assumptions, differences in responses from UNM and NMSU faculty members, and instrument psychometric properties). Incomplete or missing values were replaced with means since the amount of missing data (1%), did not meet the Missing Completely at Random (MCAR) criterion (i.e., Little's MCAR test) that is considered sufficient evidence to delete missing values listwise when less than 5% is missing (SPSS Inc., 2010). George and Mallery (2007) note that a rule of thumb is the acceptability of replacing up to 15% of missing data with the mean of the distribution with "little damage to the resulting outcomes" (p. 48).

Item analysis. Item analysis consisted of an examination of the 42 survey item means, standard deviations, inter-item correlations, and item-total correlations. Descriptive data included means, standard deviations, and frequencies of each item based on the five-point Likert scale. The highest mean score was Item 2, "Promotes institutional interests rather than self interest," and the lowest mean score was Item 25, "Increases student enrollment." Using Nunnally's (1978) guidelines for acceptable item-total correlations being near or greater than .20 and less than .70, all 42 of the items had sufficient correlations that range from $r = .16$ to $r = .66$. Cronbach's alpha = .939 for the 42-item scale. Cronbach's alpha would increase to .940 if the individual items "Increases

faculty quality,” “Promotes institutional interests rather than self interest,” “Decreases student to faculty ratio,” and “Increases student quality” were deleted. Given the results of the item-total correlations and Cronbach’s alphas with items deleted, no items were eliminated from the scale due to lack of homogeneity of the construct (Leske, 1991) or to increase internal consistency if items were deleted.

Factor analysis. PCA was performed to 1) determine the number of factors derived from the items, 2) identify naming conventions for the factors, 3) refine the theoretical and conceptual model that includes key variables related to president assessment criteria, and 4) identify opportunities to reduce the number of items in an actual president assessment instrument. Examining the inter-item correlation matrix for the 39 items included in the final PCA, 1377 of the total 1482 correlations (93%) were significant at $p = .05$. Three of the 42 items (Item 25, “Increases student enrollment,” Item 26, “Displays courage when faced with the challenges of university governance,” and Item 38, “Increases the success of students”) were eliminated from the final analysis because they appeared to be “trivial factors.” Gorsuch (1983) identifies trivial factors as those that do not have at least two of three loadings above a predetermined level (e.g., .30 for a minimum $N = 175$) or those factors without a unique set of defining variables. Following Gorsuch’s guidelines, PCA was repeated only two additional times after the initial factor solution to eliminate trivial factors.

Using $KMO = .908$ as one measure of sampling adequacy (MSA) the sample size is considered “superb” (Field, 2005, p. 650). The anti-image correlation matrix was used for additional evidence of MSA. Referring to Field’s recommendation that the diagonal elements of the anti-image correlation matrix should be greater than $r = .5$, the values for

this analysis ranged from $r = .693$ to $r = .961$ and relatively small correlations off the diagonal elements indicating that the sample size was adequate. The KMO and anti-image correlation matrix tests indicated that the correlations among the individual items were strong enough to suggest the correlation matrix was factorable (Pett, Lackey, and Sullivan, 2003). One final MSA criterion is the ratio of the number of subjects included in the factor analysis to the number of variables (items) being measured. Various authors recommend there should be between five and ten subjects per variable (Field, 2005; Kass & Tinsley, 1979; Nunnally, 1978). For this factor analysis, the number of subjects to items was 6.67 (280/42). An additional indication that factor analysis was an appropriate statistical test was Bartlett's test of sphericity. Bartlett's test of sphericity ($\chi^2 = 4121.72$, $p = .000$) provided evidence that the R -matrix was not an identity matrix and that the data were acceptable for factor analysis (George & Mallery, 2007).

Factor extraction. PCA with Varimax (orthogonal/uncorrelated) and Promax (oblique/correlated) rotations was used to explore the constructs and related dimensions associated with president assessment criteria. Using Kaiser's (1960) criterion of retaining only those factors with eigenvalues exceeding 1.0, there were eight factors retained in the factor analysis of final survey data. The Scree plot revealed between a five- to eight-factor solution; however, the plot was difficult to interpret due to the small differences in eigenvalues after factor three. The reproduced correlation matrix contains the computed residuals (differences) between observed and reproduced correlations as an indicator of the factor model fit. Using Field's (2005) guideline of 50% as the maximum, the non-redundant residuals equal or greater than $|.5|$ for the final survey data set was only 29% indicating there was an adequate model fit for factor analysis. Another indicator of

goodness of fit for the factor solution is having small ($r \leq .1$) reproduced residual correlations (Leske, 1991; Tabachnick & Fidell, 2007). For this factor extraction, only 27/1482 (2%) of the residual correlations exceeded $r = .1$ as further evidence of a tenable factor solution.

Factor rotation. The following criteria were used to analyze eight-, nine-, and ten-factor solutions using Varimax rotations: 1) items had substantive loadings equal or greater than $|.3|$ (Gorsuch, 1983; Stevens, 2002) and 2) there were at least two items with substantive loadings (Gorsuch, 1983). The nine- and ten-factor solutions contained trivial factors as defined by Gorsuch (1983) so the three variables related to these trivial factors were eliminated to clarify the factor structure that resulted in an eight-factor solution. Since Cronbach's alpha was relatively high (.935), a Promax rotation was performed to determine if the factors were highly correlated (Leske, 1991). The component correlation matrix reveal there were five medium ($r \geq .3$) to large ($r \geq .5$) correlations (Cohen, 1988) between Factor 1 through Factor 6 and several small ($r \leq .1$) to medium correlations between Factor 2 through Factor 7. The relatively high correlations among the factors indicated that a Promax rotation was more appropriate than a Varimax rotation. However, the pattern matrix computed from the Promax rotation was identical to the rotated component matrix computed from the Varimax rotation with the exception of one item. For the sake of simplicity and ease of interpretation (Pett, Lackey, & Sullivan, 2003; Tabachnick & Fidell, 2007) and since the oblique Promax rotation did not provide greater insight into the factor structure, the findings from the orthogonal Varimax rotation was considered sufficient analysis of the final survey data. The total variance explained by the eight-factor solution was 56.41%.

An eight-factor solution was consistent with the rule of thumb that the expected number of factors should be between one-fifth and one-third of the total number of variables (items) in the factor analysis (Gorsuch, 1983). Given the 39 variables included in the final factor extraction, there should have been 8 to 13 factors in the factor solution.

Factor interpretation and naming. After evaluating and refining the factors, the researcher developed preliminary names for the factors based on preliminary and intermediate analysis findings and another review of literature that defined leadership and management characteristics. Table L-1 in Appendix L contains the naming conventions and the corresponding factor loadings from the rotated factor structure based on the responses to items in the final survey ($N = 280$). After developing preliminary names for the factors, the research conducted a focus group interview and two individual interviews to obtain feedback on the naming conventions.

The final names selected for the eight factors are as follows: 1) strategic leadership, 2) consideration, 3) continuous improvement, 4) university mission support, 5) interpersonal competence, 6) stewardship, 7) academic quality, and 8) responsibility. Factor 1, *Strategic Leadership*, included 14 items typically associated with senior leadership in an organization including 1) clearing obstacles that stand in the way of positive change, 2) developing realistic goals and objectives to implement strategy, 3) providing clear direction and expectations, 4) making sound decisions based on benefit-risk analysis, 5) monitoring progress in achieving stated goals and objectives, 6) providing support to other institutional leaders, 7) building cooperative teams, and 8) rewarding superior performance. Factor 2, *Consideration* (Bass, 2008), contained three items related to displaying professional courtesy, openly sharing information with

others, and demonstrating integrity. Factor 3, *Continuous Improvement*, incorporated five items: 1) articulating the university story, 2) promoting statewide education initiatives, 3) providing a framework that aligns strategy and people to the university mission, 4) incorporating lessons learned into planning and operations, and 5) benchmarking with other institutions to identify improvement opportunities. Factor 4, *University Mission Support*, included the following four items: 1) securing adequate resources to support the university mission, 2) advocating for the university to strengthen its position and reputation, 4) recruiting high-quality people, and 5) displaying a high degree of job competence as the senior executive leader. Factor 5, *Interpersonal Competence* (Bass, 2008), contained four items: 1) serving as an inspiration to constituents and stakeholders, 2) building trusting relationships with constituents and stakeholders, 3) recruiting and retaining employees who reflect the diversity of the State, and 4) promoting harmony among constituents and stakeholders. Factor 6, *Stewardship* (Gruder, 2008), was comprised of four items: 1) showing appreciation to others for support of the university, 2) making informed decisions based on best available information, 3) maintaining awareness of the climate of the institution, and 4) making responsible decisions regarding the allocation of resources. Factor 7, *Academic Quality*, included the following three items: 1) decreasing the student/faculty ratio, 2) increasing the quality of faculty, and 3) increasing the quality of incoming students. Finally, Factor 8, *Responsibility* (Bass, 2008), included two items: 1) promoting institutional interest and 2) making judicious decisions regarding the selection of senior administrators.

Factor scales. To increase the interpretability and utility factors, factor-based scales were developed by summing the scores for only those items that were included for

a given factor (Pett, Lackey, & Sullivan, 2003). Table 12 contains the number participants, minimum values, maximum values, means, and standard deviations for each of the factor scales. Comparing the factor scales, *Responsibility* had the highest mean score and least amount of variance and *Interpersonal Competence* had the lowest mean score and highest variance.

Table 12. Factor Scales

<i>Factor Name</i>	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Standard Deviation</i>
Responsibility	279	2.00	5.00	4.67	.48
University Mission Support	274	2.75	5.00	4.41	.54
Stewardship	275	2.50	5.00	4.38	.52
Consideration	277	1.33	5.00	4.37	.64
Strategic Leadership	258	1.43	5.00	4.02	.58
Continuous Improvement	266	1.40	5.00	3.88	.66
Academic Quality	276	1.67	5.00	3.71	.68
Interpersonal Competence	274	1.25	5.00	3.67	.73

Examination of factor interdependence and reliability. The factor-based scales were used to examine the intercorrelations and reliabilities of the factors. The intercorrelations among the factors provide a measure of the interdependence of the factors. Cronbach's alpha for the factor scales provides a measure of the internal consistency of the items included in each of the factors. Table 13 provides a summary of the intercorrelations of the factors and the corresponding reliability coefficients with the Cronbach's alpha of .841 for all items included in the factor scales.

Table 13 reveals there were large correlations between the strategic leadership factor and the other factors with the exception of academic quality and responsibility.

Additionally, there were several medium to large correlations between other factors with the exception of a few small to medium correlations with academic quality and responsibility. Examining Cronbach’s alpha, .841 for the items included in the factor scales is “good” in terms of internal consistency (George & Mallery, 2007). Nunnally (1978) suggests that modest reliabilities of .70 are sufficient for an instrument in the early stages of research. Kline (1999) contends that values below .70 can be expected because of the diversity of constructs being measured in psychological research.

Table 13. Factor Correlations and Factor Alpha Coefficients

<i>Factor Name</i>	<i>Strat. Ldrshp</i>	<i>Consd.</i>	<i>Cont. Imp.</i>	<i>Msn. Spt.</i>	<i>Int. Comp.</i>	<i>Stew.</i>	<i>Acad. Qual.</i>	<i>Resp.</i>	<i>Alpha Coeff.</i>
Strategic Leadership	1								.906
Consideration	.497	1							.727
Continuous Improvement	.699	.386	1						.773
University Mission Support	.550	.330	.538	1					.665
Interpersonal Competence	.585	.521	.533	.369	1				.653
Stewardship	.615	.499	.523	.447	.486	1			.662
Academic Quality	.280	.213	.273	.354	.264	.294	1		.508
Responsibility	.232	.317	.247	.240	.223	.311	.163	1	.456

Examining the internal consistency of the individual factors, university mission support, interpersonal competence, and stewardship have “questionable” reliability (with 4 items in each factor scale); academic quality has “poor” reliability (with 3 items in the factor scale; and responsibility has “unacceptable” reliability (with 2 items in the factor scale) (George & Mallery, 2007). Addressing the issue of low reliability coefficients, Field (2005) argues since Cronbach’s alpha is very dependent on the number of items in the scale, relatively low numbers are not conclusive evidence that the scales are

unreliable. Conversely, high numbers are not conclusive evidence that the scales are reliable. In the case of the factor scales in this study, it is possible that the large Cronbach's alpha for strategic leadership (.906) is overestimated due the relatively large number of items and the coefficients for the factors with fewer items are underestimated.

Examination of means scores. The following tables provide the mean scores for the items included in the final survey related to: 1) assessment instrument content (i.e., candidate assessment criteria), 2) external factors that could potentially impact the ratings derived from an assessment instrument, and 3) format of an assessment instrument and its application (i.e., formality, purpose, structure, participants, and frequency).

Table 14. Top 25 Scores for Candidate Assessment Instrument Content

<i>Items</i>	<i>Mean Scores</i>
Promotes institutional interests	4.75
Makes responsible resource allocation decisions	4.74
Demonstrates integrity	4.67
Displays job competence	4.61
Makes judicious decisions on administrator selection	4.59
Secures resources to support university mission	4.47
Maintains awareness of institutional climate	4.44
Makes informed decisions	4.44
Maintains effective control of the budget	4.33
Recruits high caliber people	4.32
Increases faculty quality	4.29
Encourages open sharing of information	4.26
Advocates for the university to strengthen higher education	4.24
Promotes positive change	4.22
Develops realistic plans to implement strategy	4.21
Displays professional courtesy	4.19
Establishes realistic goals and objectives	4.19
Builds trusting relationships with constituents/stakeholders	4.17
Adapts to changes	4.14
Provides clear directions and expectations	4.12
Promotes statewide education initiatives	4.02
Promotes negotiation to resolve conflicts	4.00
Clears obstacles to progress	4.00
Provides thoughtful responses to constituents/stakeholders	3.99
Supports institutional leaders	3.97

The Likert scale definitions for responses to the items were as follows: 5 = “critically important,” 4 = “very important,” 3 = “important,” 2 = “slightly important,” and 1 = “not important.” For the purpose of analysis, the following ranges are used to categorize the important of the items: “critically important” (4.60 – 5.00), “very important” (3.60 – 4.59), “important” (2.60 – 3.59), “slightly important” (1.60 – 2.59), and “not important” (1.00 – 1.59).

Table 14 contains the top 25 scores for assessment instrument content. These scores were derived from the items in the final survey in which faculty were asked to rate the importance of various traits, behaviors, and performance outcomes as president performance assessment criteria. There were 42 closed-ended items in the final survey related to assessment instrument content. Of these 42 items, the faculty rated promoting institutional interests, making responsible resource allocation decisions, demonstrating integrity, and displaying job competence as “critically important” assessment criteria. While still considered “very important” assessment criteria, promoting negotiations to resolve conflicts, clearing obstacles to progress, providing thoughtful responses to constituents and stakeholders, and supporting institutional leaders were at the bottom of the top 25 list of candidate assessment criteria.

Table 15 contains the highest to lowest mean scores for external factors that faculty believed could potentially influence president performance ratings. The final survey included seven closed-ended items that addressed external factors. Table 15 indicates that faculty perceived the political pressure was the top external factor that could have “very important” impact on president performance ratings. While rater errors

were “important” considerations, this factor had the lowest mean score of the seven external factor items included in the final survey.

Table 15. External Factors that Could Impact President Performance Ratings

<i>Items</i>	<i>Mean Scores</i>
Political pressures from government officials	3.92
Conflicting priorities with those in shared governance	3.69
Economic conditions	3.58
Rater knowledge of president actions/decisions	3.58
Rater knowledge of president roles/responsibilities	3.53
Scholarly criticism/skepticism	3.34
Rater errors (halo effect, leniency, central tendency)	3.07

Table 16 contains the highest to lowest mean scores for items in the final survey that addressed president assessment format. The final survey included seven closed-ended items pertaining to the format of an assessment instrument and its application. Referring to Table 16, faculty members believed that involving multiple raters, performing assessments on a regular basis, conducting formal assessments, including specific criteria in an assessment instrument, and providing assessment feedback for the purpose of personal development and improvement were “very important” characteristics of an assessment. While the score for informal assessment qualifies as a “slightly important” characteristic, this score implies that some respondents believe informal assessment is a preferable format compared to formal assessment.

Table 16. Assessment Format

<i>Items</i>	<i>Mean Scores</i>
Involves multiple raters	4.11
Includes assessment on a regular basis	3.95
Consists of a formal assessment with written policies, procedures, and assessment form	3.85
Incorporates specific assessment criteria	3.85
Provides feedback for personal development/improvement	3.85
Ties performance to compensation	3.58
Consists of an informal assessment without written policies, procedures, and assessment form	1.93

Final Survey Qualitative Findings

Final survey participants provided 90 responses to the open-ended question: “Do you have additional comments on the criteria for assessing university president success, best practices for assessing university presidents, factors that should be taken into account when assessing university presidents, and/or this survey?” In general, the responses to this question centered around the following themes: 1) high priority of academic and education quality, 2) concerns about excessive compensation for senior administrators, 3) overemphasis on athletics at the expense of academics, 4) importance of participative management, 5) standing up to political pressures, 6) importance of having other constituents and stakeholders involved in the assessment process and not just members of the board of regents, and 7) suggestions for improving survey questions. Tables K-1 through K-3 in Appendix contain the comments from final survey in an abbreviated format. Table K-4 includes the counts for the initial codes derived from survey comments using Atlas *TI*[®].

Table K-1 contains grounded theory initial, axial, and theoretical coding for recommended performance assessment criteria. The items in the first column of this table are the initial codes derived from the specific response from a final survey participant on recommended performance criterion not measured in the survey. The second column reflects the assigned axial code for the subcategory or dimension related to the initial code for each of the 49 responses. The last column is the corresponding theoretical code for each response. The theoretical code represents a “category” in grounded theory terminology that is equivalent to a “construct” using theoretical/conceptual terminology and a “factor”/“component” in quantitative analysis

terminology. In most cases, the theoretical codes correspond to the computed PCA eight-factor solution (i.e., strategic leadership, consideration, continuous improvement, university mission support, interpersonal competence, stewardship, academic quality, and responsibility). Five responses did not appear to map well with the eight-factor solution. After conducting a literature review, the following constructs were identified to account for the responses that were not compatible with factors derived from factor analysis of the quantitative data collected from the final survey: 1) *Communication* (Blanchard & Associates, 2007; Drucker, 2008), 2) *Self-mastery* (Goleman, 1998), 3) *Courage* (Treasurer, 2008), *Legitimization* (Bass, 2008), and *Intelligence* (Gardner, 2004; Goleman, 1998, 2006; Gruder, 2008).

Final Interview Findings

The purpose of the final interviews was to review the results of the study and to obtain additional validity evidence. Feedback from one focus group interview involving tenured faculty members and two individual interviews with non-tenured faculty members was incorporated into the final results of the study. The final survey participants provided the following comments:

- Rather than using the only the means of the factor scales, the researcher should investigate the standard deviations of the individual items included in each factor to gain a better sense of the relative agreement on the importance of various criteria.
- The researcher should review the preliminary naming conventions for Factor 1 (originally referred to as “Strategic Leadership/Management”), Factor 2 (originally referred to as “Character”), Factor 5 (originally referred to as “Social

Skills”), and Factor 8 (originally referred to as “Accountability”). After additional research, these factors were renamed as follows: Factor 1 “Strategic Leadership,” Factor 2 “Consideration,” Factor 5 “Interpersonal Competence,” and Factor 8 “Responsibility.”

Based on feedback from the final interviews, there was a unanimous consensus among the final survey participants that the findings and conclusions from this study were tenable and could support development of a president assessment instrument.

Integration of Qualitative and Quantitative Findings from Final Analysis

Tables 17-20 integrated the qualitative and quantitative findings derived from preliminary, intermediate, and final analysis. These tables condense the large number of items investigated during this study down to the top 25 items in each the three major categories: 1) assessment instrument content, 2) external factors, and 3) assessment format. Under the category of assessment instrument contents, 160 traits and behaviors and 62 performance outcomes were identified during the various analysis phases. Similarly, a total of 35 external factors and 46 assessment format characteristics were identified during three analysis phases.

The following criteria were used to determine which items were included in the top 25: 1) the item was measured in the final survey, 2) the item was measured in the pilot survey, or 3) the final survey or pilot survey contained a specific comment that applied to one of the three major categories. The tables also indicate items that were supported by literature reviewed for this study with a single-check (√) indicating there was only one substantiating source document and a double-check (√√) indicating there were multiple substantiating source documents.

The findings in Table 17 suggest that participating faculty believe that promoting institutional interests, making responsible resource allocation decisions, integrity, job competence, and making judicious decision on administrator selection are “critically important” (mean scores between 4.50 and 5.00) traits and behaviors that could serve as useful president assessment criteria. The relatively high means scores for these items on the final survey and final survey along with interview comments and literature reinforce the importance these items.

Table 17. Top 25 President Assessment Traits and Behaviors Criteria

Criteria	Final Survey Means	Final Survey Comments	Pilot Survey Means	Pilot Survey Comments	Interview Comments	Literature Review
Promotes institutional interests	4.75	√	4.71	√	√	√
Makes responsible resource allocation decisions	4.74	√	4.15	√	√	√
Demonstrates integrity	4.67		4.75		√	√√
Displays job competence	4.61		4.56		√	√√
Makes judicious decisions on administrator selection	4.59	√			√	√
Maintains awareness of institutional climate	4.44				√	√
Makes informed decisions	4.44		4.00			√
Maintains effective control of the budget	4.33				√	√
Recruits high caliber people	4.32	√	4.29			√
Encourages open sharing of information	4.26	√	4.58		√	√√
Advocates for the university to strengthen higher education	4.24	√	4.24		√	√
Promotes positive change	4.22	√	3.98			√
Develops realistic plans to implement strategy	4.21		4.18	√		√
Displays professional courtesy	4.19		4.33			√√
Establishes realistic goals and objectives	4.19	√	4.18			√
Builds trusting relationships with constituents/stakeholders	4.17	√	4.63			√√
Adapts to changes	4.14	√	4.35			√
Provides clear directions and expectations	4.12	√	3.68			
Promotes statewide education initiatives	4.02	√			√	√
Promotes negotiation to resolve conflicts	4.00		3.95		√	√
Clears obstacles to progress	4.00		4.33			√
Provides thoughtful responses to constituents/stakeholders	3.99	√			√	√
Supports institutional leaders	3.97	√	4.29			√
Establishes effective teams	3.95				√	√√
Promotes harmony	3.91	√	3.81			√

While items such as providing thoughtful responses to constituents/stakeholders, supporting institutional leaders, establishing effective teams, and promoting harmony round out the bottom of the list, the participants still considered these items as “very important” president assessment criteria and interviews and literature corroborate the final survey results.

Referring to Table 18, with the exception of the top two items (securing resources to support the university mission and increasing faculty quality), there appeared to be a relatively distinct difference between the final survey scores on items related to traits and behaviors and on items related to performance outcomes. This result was consistent with preliminary and intermediate analysis and was substantiated by authors who expressed reservations about the ability of a university president to have a direct impact on organizational level performance outcomes (Coens & Jenkins, 2002; Illgen & Favero, 1985; Yudof & Busch-Vishniac, 1996).

Despite the differences in scores and the skepticism surrounding a president’s impact, faculty participants rated many performance outcomes as being “important” to “very important” as evidenced in Table 18. For example, the pilot survey contained 14 items that respondents scored as being “very important” that were not addressed in the final survey in order to limit its length. There is also a substantial amount of literature that reinforces the importance of performance outcomes; however, authors other than Frank Rhodes (2001), President Emeritus of Cornell University, do not emphasize the direct relationship between president performance and organizational level performance outcomes. On the other hand, there is substantial evidence that university presidents are held responsible by their boards for achieving organizational goals and objectives that

implies that performance outcomes are relevant assessment criteria (Arizona Board of Regents, 2005; Engelkemeyer, 1999, 2008; Nason, 1997; University of Alabama System, 2003; University of Nevada Las Vegas, 2003; University of New Mexico, 2007b, 2008, 2009; University of Washington Board of Regents, 2006). Given the perceived importance of performance outcomes, a separate table of the top 25 performance outcomes is included in this study to ensure these items were not overshadowed by assessment criteria associated traits and behaviors that generally had higher mean scores compared to performance outcomes.

Table 18. Top 25 President Assessment Performance Outcome Criteria

Items	Final Survey Means	Final Survey Comments	Pilot Survey Means	Pilot Survey Comments	Interview Comments	Literature Review
Secures resources to support university mission	4.47	√		√	√	√√
Increases faculty quality	4.29				√	√
Increases student success	3.72					√
Reduces student/faculty ratio	3.44		4.29		√	√√
Increases incoming student quality	3.39				√	√√
Recruits/retains employees that reflect diversity of the state	3.26					√
Increases student enrollment	2.46		3.17			√
Increases % of full time faculty			3.97			
Increases stakeholder satisfaction			3.96		√	√
Increases student graduation rates			3.96			√√
Improves campus quality of life			3.94			√
Reduces average number of students per class			3.92			√
Increases level of faculty compensation			3.89			√
Increases student retention rate			3.89			√
Increases % classes taught by tenured faculty			3.87			√
Investment in facility modernization			3.82			√
Increases amount of foundation gifts			3.80			√
Increases ratings by peer institutions			3.77			√√
Increases amount of grant funding			3.69			√
Increases involvement in local community			3.69			√
Increases amount of scholarships for students			3.60			√
Improves placement of graduating students			3.53			√
Increases number of scholarly publications by faculty			3.51			√
Increases % of faculty with top terminal degrees			3.51			√
Increases investment in advanced info technologies			3.47			√

Examining the rank ordering, there were 18 items that faculty participants believed were “very important” indicators of president success and useful performance assessment criteria based on final survey and pilot survey mean scores from 3.60 – 4.59. While the interviews reinforced the importance of only six of these items, literature substantiated the importance of several other performance outcome items.

Table 19. Top 25 External Factors Influencing Performance Assessment Ratings

Items	Final Survey Means	Final Survey Comments	Pilot Survey Means	Pilot Survey Comments	Interview Comments	Literature Review
Political pressures from government officials	3.92	√			√	√√
Conflicting priorities with those in shared governance	3.69		4.15		√	√√
Economic conditions	3.58		3.37		√	√√
Rater knowledge of president actions/decisions	3.58					√√
Rater knowledge of president roles/responsibilities	3.53					√√
Scholarly criticism/skepticism	3.34				√	√√
Rater errors (halo effect, leniency, central tendency)	3.07		3.67		√	√√
Perceptions of followers on the attributes of a good leader			3.73		√	√√
Experience of raters			3.70		√	√√
Lack of real authority of president			3.47		√	√√
Faculty resistance						√√
Extent of the president’s powerbase						√√
Degree of support from superiors						√√
Availability of resources (e.g., from the state legislature)						√√
Unrealistic job expectations						√√
Reluctance to change on parts of constituents/stakeholders						√√
Institutional inertia						√√
Instability of customer demand						√√
Legal constraints						√√
Complexity of the job						√√
Setting of unrealistic goals						√√
Lack of cooperation of constituents/stakeholders						√√
Organizational bureaucracy						√√
Cultural influences						√√
Lack of defined performance criteria						√√

Table 19 indicates there were only two items that final survey participants believed were “very important” external factors that could have an impact on president performance

ratings: political pressures from government officials and conflicting priorities with those who occupy positions in shared governance. There was a substantial amount of literature that confirmed the potential negative effect of political pressure from a multitude of sources as well as economic conditions, resistance to change, job complexity, cultural influences, and rater error to name a few.

Table 20. Top 25 Characteristics of an Effective President Assessment Format

Items	Final Survey Means	Final Survey Comments	Pilot Survey Means	Pilot Survey Comments	Interview Comments	Literature Review
Involves multiple raters	4.11	√	4.22		√	√√
Includes assessment on a regular basis	3.95		4.04		√	√√
Consists of a formal assessment with written policies, procedures, and assessment form	3.85	√	3.80		√	√√
Incorporates specific assessment criteria	3.85		3.78		√	√√
Provides feedback for personal development/improvement	3.85				√	√√
Ties performance to compensation	3.58		3.64		√	√√
Consists of an informal assessment without written policies, procedures, and assessment form	1.93				√	√√
Includes objective performance assessment criteria			3.68		√	√√
Includes open-ended questions, qualitative criteria		√		√	√	√√
Includes closed-ended questions, quantitative criteria		√			√	√√
Includes public dissemination of results		√			√	√√
Accounts for external factors		√			√	√√
Involves vetting with experts in instrument design		√				
Maintains confidentiality of specific assessment results		√				√√
Involves training for raters		√				√√
Includes focus group and committee participation		√				√√
Includes an organizational climate assessment				√	√	√
Its purpose is not to serve as a punishment tool				√	√	√
Its purpose is to provide feedback					√	√√
Its purpose is to increase accountability					√	√√
Its purpose is to support retention decisions					√	√√
Its purpose is to set the tone for the organization					√	√
Incorporates anonymity of the raters					√	√
Requires a balance of power among constituents					√	√
Includes a self-assessment					√	√√

Table 20 contains items related to the format of an assessment instrument.

Webster's dictionary defines format as "general arrangement or plan" (Agnes & Guralnik, 2001, p. 556). In the context of this study, format pertains to: 1) formality (i.e., a formal assessment includes documented policies, procedures, assessment instrument, and report) 2) purpose of the assessment, 3) structure of the assessment instrument (e.g., closed questions, open questions, objective criteria, subjective criteria), 4) participants in the assessment, and 5) frequency at which the assessment instrument should be applied.

Referring to Table 20, final survey participants score the following items as "very important" to the assessment of a university president: 1) involves multiple raters, 2) includes assessment on a regular basis, 3) consists of a formal assessment with written policies, procedures, and an assessment form, 4) incorporates specific criteria, and provides feedback for personal development and improvement. The individual and focus group interviews along with the literature review corroborate these final survey results. The interviewees stressed the importance of 360-degree feedback that involved multiple constituents and stakeholders to include: 1) members of the board of regents, 2) faculty, 3) students, 4) staff, 5) members of faculty committees, 6) alumni, and 7) legislators. Providing feedback for self-improvement was the most frequently mentioned purpose of an assessment. Interviewees and survey respondents tended to favor formal assessment that incorporated a combination of qualitative (open-ended) items and quantitative (closed-ended) items. In terms of frequency, the interviewees generally agreed assessments should be performed on an annual basis.

Summary of Findings on Assessment Instrument Content and Format

The list of performance criteria that could make of the content of an assessment instrument for a university president is very long. This study identified 222 candidate traits, behaviors, and performance outcomes derived from theories and concepts found in literature, interviews of faculty members, a pilot survey conducted at UNM, and a final survey conducted at UNM and NMSU. Butler (2007) provides additional evidence of the extensiveness of potential assessment criteria by citing an AGB literature review that revealed over 200 measures of effectiveness, efficiency, and outcomes that colleges and universities use to assess performance outcomes alone. However, several authors (Bass, 2008; Blanchard & Associates, 2007; Finkelstein, Hambrick, & Canella, 2009; Gini, 1995) stress that leadership is situational and that each institutional context is different requiring a tailored assessment approach for a given institution at a particular moment in time (Munitz, 1978).

Applying the principle of leadership being situational and the importance of using context as a filter for designing an assessment, this study successfully reduced the vast number of potential assessment criteria to those that participating UNM and NMSU faculty members believed were most important for their institutions at a particular moment in time. The items in Table 17 and Table 18 reflect the opinions of 280 faculty members from UNM and NMSU on the most important assessment criteria that the universities should consider based on the items included in the surveys.

The factor analysis, that included many of the items in Table 17 and a few of the items in Table 18, revealed eight overarching constructs (see Table 13) to which these items relate. Since the items within a given construct correlate highly, they are

considered to be measuring the same thing (Field, 2005; Nunnally, 1978; Pett, Lackey, & Sullivan, 2003). If a university wants to limit the number of items in an assessment instrument, it could choose a representative set of items within each construct and have an added degree of confidence that the reduced set of items adequately measures the constructs of interest. Performing grounded theory analysis of performance assessment criteria provided by final survey participants, five additional constructs were identified (see Table K-1) to account for items that did not connect directly to the eight constructs derived from factor analysis of the quantitative data collected from the final survey.

Another area that should be considered in terms of the context of an assessment is external factors that can bias or influence the ratings a president could receive on various performance criteria. Various authors point to a myriad of external factors that can affect performance assessment of individuals including university presidents (Armstrong, 2009; Bass, 2008; Berk, 1986; Bernardin & Beatty, 1984; Duderstadt & Womack, 2003; Finkelstein, Hambrick, & Cannella, 2009; Grote, 2002; Illgen & Favero, 1985; Latham & Wexley, 1981; Rhodes, 2001). Table 19 provides a summary of the external factors investigated during this study and the relative importance of some of these factors based on the feedback from the participating faculty members from UNM and NMSU. While these external factors do not directly address the research question for this study, they may be important considerations in the application of an assessment instrument.

Turning to the final major area of exploration, Table 20 contains a list of items that address the format for a university president assessment. Again, participating UNM and NMSU faculty members rated the importance of some of these items in performing an effective assessment of a university president. A substantial amount of literature

substantiates these findings on assessment formats; however, the literature review, pilot survey, interviews, and final survey revealed that some have reservations about the utility of individual performance assessment in general (Coens & Jenkins, 2002) and formal president assessment in particular (Kauffman, 1978; Ingram & Weary, 2000). Despite these reservations, president performance assessment continues to be a widespread practice in universities (Schwartz, 1998) and there are optional formats that universities can consider in designing their assessment processes and assessment instruments.

Findings from Reliability and Validity Assessment

The paragraphs that follow contain evidence of the reliability and validity of qualitative and quantitative data collected during this study. The first major paragraph and subparagraphs discuss the dependability and validity of qualitative data from interviews with dependability being synonymous with the term reliability in quantitative analysis. The second major paragraph and subparagraphs include findings on the reliability and validity of quantitative data from the pilot survey and final survey. The final major paragraph highlights the findings on reliability and validity from a mixed-methods perspective.

Qualitative Dependability and Validity Findings

The individual interviews and focus group interviews were the primary sources of data for qualitative analysis conducted during this study. The open-ended questions allowing respondents to provide comments on the pilot survey and final survey were an additional source of qualitative data. The paragraphs that follow address the dependability, credibility, transferability, and confirmability of qualitative data.

Dependability

There was sufficient evidence of the consistency of data across subjects. For the initial individual and focus group interview, there was considerable agreement among the participants on the importance of various traits, behaviors, and performance outcomes that could serve as a basis for determining the content of a president assessment instrument. The responses to initial interviews were generally consistent in the areas of external factors affecting performance ratings and the format of an assessment instrument. While the initial interviewees expressed different concerns and considerations with respect to the main areas of interest in this study, there were no fundamental inconsistencies in their viewpoints. The overlap in the findings from the interviews, pilot survey, and final survey provided further evidence of the dependability of qualitative data. Finally, participants in the second and third round of interviews agreed that the findings from the interviews and surveys appeared to be consistent.

Credibility

The outcomes from triangulation, peer debriefing, negative case analysis, referential adequacy checks, and member checks served as sources of evidence for confirming credibility of qualitative data. This study involved multiple data sources, methods, and theoretical schemes as part of the triangulation process to ensure data credibility. The triangulation process revealed consistencies in the data and resolved differing viewpoints of the study participants. Members of the researcher's committee and participants in the follow-up interviews provided positive feedback on the substantive, methodological, and ethical aspects of the study as part of the triangulation process. Addressing the substantive aspect of triangulation, individuals who had not been

involved in previous interviews agreed that the findings related to the content and format of a president assessment instrument were understandable and credible. Additionally, feedback from peer debriefings reinforced the methodological approach used in this study.

One issue that came up during a peer debriefing early in the study was the sensitivity of the study topic. In response to this concern, additional steps were taken to ensure confidentiality. These steps included restricting participation in focus group interviews to tenured faculty members and allowing non-tenured faculty members to participate only in private, individual interviews. As an additional protective measure, the researcher did not reveal the names or specific positions of any faculty members who participated in interviews. Instead, the researcher only revealed the colleges/schools to which the interview participants were assigned and as well as the general expertise they had that was relevant to this study.

The researcher conducted a negative case analysis by identifying and resolving differing perspectives among study participants. For example, the researcher had the initial impression that two interviewees were opposed to formal assessment of a president. However, after analyzing their comments in more detail during the intermediate analysis phase, the two interviewees suggested conditions under which formal assessment could part of a viable format.

Throughout the study, the researcher performed referential adequacy checking by revisiting data collected from different participants at different points in the AIDA model to check for consistency in the findings. The researcher performed member checks through formal, informal, and continuous reviews of data and emerging results

throughout the study to obtain further evidence of data credibility. Using an approach that Creswell (2007) refers to as a *data analysis spiral*, the researcher performed data collection, data analysis, and results preparation as an integrated and recursive activity during this study. The data analysis spiral, an integral part of the AIDA model, is consistent with an exploratory research design that involves “learning by doing” (Dey, 1993, p.6) and “pulling out threads of stories we discover in data” (Dey, 1995, p.78) as events unfold rather than waiting to analyze data toward the end of a study.

Transferability

To provide evidence of the transferability of the findings to other subjects and contexts, the researcher wrote *thick descriptions* of all interview recordings in grounded theory coding worksheets. Entries in the worksheets included *in vivo codes* (Creswell, 2007) which are the exact words of the participants. The researcher used these *in vivo* codes during the initial, axial, and theoretical coding processes as part of the grounded theory research design incorporated in this study.

The initial target population for the study was limited to the faculty at UNM. In order to increase transferability of the study results to other universities, the researcher expanded the target population to include the faculty at the NMSU. However, since the qualitative data collection was limited to UNM, there was insufficient evidence of transferability of the results to institutions other than UNM. The researcher used purposive sampling of faculty members instead of random sampling for the selection of interview participants, which limited representativeness of the participants and transferability of the data.

Confirmability

As evidence of confirmability, the researcher took steps to control personal motivations, interests, and perspectives that could have potentially biased qualitative data. For example, the researcher transcribed audio recordings from interviews directly onto ground theory coding worksheets without interpreting or paraphrasing interviewee statements to add or to clarify the meaning of the statements. During the analysis and reporting process, the researcher retained the original words provided by the interviewees instead combining them with synonyms for the purpose of expediency. Another step that supported the confirmability was the use of a pilot survey and final survey that consisted almost exclusively of closed-ended questions rather than open-ended questions that usually require interpretation and translation thereby reducing the objectivity of the data.

As a final step, the researcher kept a reflexive journal of personal perceptions, meanings, and contexts to increase awareness of those factors that could have biased qualitative data collection, analysis, and reporting of study findings and results. The following researcher experiences could have biased study findings, conclusions, and recommendations:

- All previous experience in the workforce (over 35 years) involved formal assessment including an instrument with quantitative and qualitative assessment criteria. Researcher's familiarity with a formal assessment system and lack of experience with informal assessment could have biased data collection and analysis.
- In some previous work experiences, performance assessment had a direct impact on promotion and compensation decisions and was a useful tool for individual

development and improvement. In other cases, performance assessment was more of a cursory process with limited utility for the supervisor, individual, or the organization because of inflated ratings and lack of interest. Researcher's preference of having a more detailed assessment system that includes performance incentives and focuses on performance improvement could have influenced data collection and analysis.

In preparation for the dissertation hearing, the researcher informed dissertation committee members of personal biases that could have negatively affected the objectivity of the study. During the dissertation, the committee members confirmed that the methodology, findings, and results of the study appeared to be free of researcher bias.

Quantitative Reliability and Validity Findings

The pilot survey and final survey were the primary sources of data for quantitative analysis conducted during this study. In some cases quantitative analysis provided further evidence of reliability and validity of the data – in other cases, it did not. The paragraphs that follow address reliability, statistical conclusion validity, internal validity, content validity, construct validity, and external validity of quantitative data. Criterion validity is not addressed since the final survey was a cross-sectional survey and did not involve its application at different points in time (to check predictive aspect of criterion validity) or to different target populations (to check concurrent aspect of criterion validity). Consequential validity is not addressed since the study did not involve the actual performance of the president and additional steps were taken to ensure confidentiality of the survey participants.

Reliability

With the exception of the low coefficients for some of the factor scales with relatively few items (see Table 13), Cronbach's alpha for the pilot survey (.960), final survey (.943), factor analysis items (.935), and consolidated factor scales (.841) were relatively high. It is likely that the internal consistency coefficient was overestimated for the pilot survey because of the large number of items (111 total). It is also possible that the coefficients for the factors scales were underestimated because of the small number of items (only 2 to 4 items for the factor scales with Cronbach alphas > .7).

Statistical conclusion validity

There was positive evidence of statistical conclusion validity using the following criteria: 1) reliability of the scores, 2) tenability of assumptions, 3) statistical power and 4) control Type I error for the statistical tests (Shadish, Cook, & Campbell, 2002). Evidence of reliability is presented in the preceding paragraph.

The following results provide evidence that the assumptions for PCA were met:

- 1) KMO and the anti-image correlation matrix confirmed excellent sample size,
- 2) Bartlett's χ^2 test indicated that the correlation matrix was not an identity matrix, and
- 3) the factor scales had very few outliers (seven factors had less than 1% outliers and one factor [responsibility] had 2.5% outliers). Multivariate normality was not assessed for PCA since it is not a requirement (Tabachnick & Fidell, 2007).

The following results provide evidence that the assumptions for MANOVA were met: Box's *M* test indicated equality of the covariance matrices for dependent variables (DVs) and Levene's test for homogeneity of variance indicated equality of error variances for DVs. The assumption of multivariate normality did not appear to be met

for MANOVA. Since SPSS[®]/PASW[®] do not include a diagnostic tool for analysis of multivariate normality, Field (2005) suggests that inspection of univariate normality is a practical substitute. The Kolmogorov-Smirnov test of the factor scales indicated deviations from univariate normality for all eight factors. Inspections of histograms revealed relatively large negative skewing for four factors (consideration, university mission support, stewardship, and responsibility) and Q-Q plots reinforced deviations in normality for these four factors. Despite these indications of the lack of multivariate normality, Stevens (2002) contends that MANOVA is “robust with respect to Type I error against non-normality” (p. 263). Stevens also suggests that non-normality can affect the outcome of Box’s test giving a false indication of the equality of covariance matrices; however, this was not the case in this analysis.

Statistical power was only applicable to MANOVA that was applied to determine if there was a statistically significant difference between the UNM and NMSU scores for evidence of external validity. Power $(1-\beta) = .90$ for the MANOVA test of the factor scales that included 39 performance assessment criteria. This computed power translates to a 10% probability of a Type II error (incorrectly failing to reject the null hypothesis when it is false). In this analysis, Type I error was controlled by conducting MANOVA rather than separate ANOVAs in which multiple tests on the same data increases the familywise error rate (Field, 2005).

Internal validity

Use of a cross-sectional survey reduced maturation and instrumentation threat to internal validity. On the other hand, the high non-response rate for the final survey posed a significant threat to internal validity because there is a significant possibility that

relationship among the variables (items) in the final survey could have been influenced by factors such as age, faculty status, teaching experience, and general interest in the topic for this study.

Content validity

The following techniques used in this study contributed to the content validity of the quantitative results of the final survey: 1) performing a comprehensive literature review, 2) using Critical Incident Technique (CIT) to identify tasks associated with university president performance, 3) involving faculty members in the study with relevant subject matter expertise, 4) ensuring the instrument had a representative collection of items, and 5) using a methodical approach to developing the final survey instrument.

Construct validity

Using authoritative literature and feedback from interviewees who had substantial knowledge and experience with senior leadership and performance assessment helped operationalize the primary domains of interest and supported construct validity. However, the relatively high degree of correlation between six of eight factor scales (see Table 13) provided contradictory evidence of *discriminant validity*, which would normally bring into question the operationalism of the constructs. Given the relatively high correlations among most of the factor scales, there is substantial evidence that these factors are closely related to the central theme of strategic leadership. Turning to the question of *convergent validity*, there was positive evidence of this aspect of convergent validity with all factors with the correlations of items aligned with a particular factor scale being higher than correlations with items aligned with different factor scales. The factor scales

for academic quality and responsibility provided evidence of a moderate degree of discriminant and convergent validity as part of construct validity.

External validity

Since the final survey did not involve random sampling, the external validity of the quantitative data is questionable. A comparison of the demographic backgrounds of survey participants at UNM and the demographics of the UNM faculty (University of New Mexico, 2010) indicated overrepresentation of more experienced faculty members and underrepresentation of minority faculty members. MANOVA was conducted to investigate differences between UNM and NMSU faculty members in their perspectives of the importance of various president assessment criteria for additional evidence of external validity.

The omnibus test of the main effect of university ($N = 274$) was statistically significant, Wilk's $\lambda(8,222) = 2.46, p = .014$, partial $\eta^2 = .069$, $1-\beta = .90$. The one-degree-of-freedom between subjects test for NMSU faculty ($M = 3.50, SD = .67$) and UNM faculty ($M = 3.80, SD = .65$) for perceived importance of academic quality was statistically significant at the specified .05 level, $F(265) = 12.25, p = .001$, partial $\eta^2 = .043$, $1-\beta = .94$, 95% CI [3.71, 3.90]. Comments from the final survey indicated that some faculty members believe that the provost has more responsibility for academic quality than the president does. Despite the significant difference in perspective of the importance of academic quality, the effect size was relatively small. Therefore, the differences in perspectives between UNM and NMSU survey participants appeared to be minimal and provided some additional evidence of external validity and lack of biases between the two groups.

Summary of Findings from Mixed-Methods Reliability and Validity Assessment

The methodology applied in this study along with the study results provide a moderate amount of reliability and validity evidence. Drawing qualitative and quantitative data from the same population, integrating the data, addressing many of the same questions, using a systematic approach to develop the final survey, and applying transformation codes (see Table A-1) to link qualitative and quantitative data contributed to mixed-methods validity. In general, Table 17 through Table 20 provide additional evidence of mixed-methods reliability and validity of the study results.

CHAPTER FIVE

DISCUSSION

Chapter 5 provides conclusions and recommendations based on the findings in Chapter 4. The major sections of the chapter are 1) assessment instrument development conclusions, 2) assessment instrument content and format conclusions, 3) implications of the study, 4) limitations of the study, 5) recommendations, and 6) summary of the discussion. The primary products in this chapter are answers to the research questions. Accordingly, Chapter 5 describes a revised AIDA model that universities should consider when developing new or revised assessment instruments for their presidents. The chapter also highlights the preferred content and format of an assessment instrument based on the perspectives of key constituencies in the State of New Mexico – the faculty at UNM and NMSU.

Another key product is a revised theoretical and conceptual framework for university president assessment. The revised framework illustrates the relationships between pertinent variables that researchers should consider in the development of a president assessment instrument. The findings in Chapter 4 served as a starting point to the formulation of hypotheses on the relationships between pertinent variables. The results of hypothesis testing of these variables in future research could provide insights into the more relevant items to include in an assessment instrument and to increase the reliability and validity of performance ratings derived from that instrument.

Assessment Instrument Development Conclusions

The paragraphs that follow address Research Question 1: *What approach can UNM and other universities use to develop an effective performance assessment*

instrument for their presidents? The answer to the first research question is derived from the findings associated with the application of the AIDA model during this study. As such, the following conclusions are based on the experiences and perceptions of the researcher and the literature available on assessment instrument development.

Conclusions on Assessment Instrument Development

There were several positive aspects of the AIDA model (see Figure 9). First, the model provided a structured approach for developing an assessment instrument using an inductive, bottom-up approach. The literature review provided a significant amount of information on theories and concepts associated with leadership and performance assessment. However, there was much less information available on assessment instrument development, particularly for university presidents. Grounded theory and Critical Incident Technique (CIT) served as useful tools for aggregating and synthesizing the vast amount of general information into a usable form and for identifying relevant items to include in the initial interview guides and the pilot survey. The initial individual interviews, focus group interviews, and pilot survey provided more specific data that corroborated, contradicted, or expanded upon data from the literature review. The follow-up interviews and final survey allowed the researcher to identify the more salient characteristics of a performance assessment instrument in terms of its content and format based on the perspective of faculty at the two largest public universities in New Mexico.

Another positive aspect of the AIDA model is that it fully integrated qualitative and quantitative research methods allowing the data to be examined continuously from multiple perspectives. Breaking data analysis into three phases (i.e., preliminary, intermediate, and final analysis) served as a building-block approach to evaluate and

refine results as well as the means to step back and look at consistencies and inconsistencies in previous analysis. The integrated and mutually supporting phases in the AIDA model provided evidence of reliability and validity of the data collected and analyzed during this study and is discussed later in this chapter in more detail.

An additional advantage of the AIDA model is that it provides provisions for a robust approach for assessment instrument development that is adaptable to needs of a particular university. For example, it includes the use of statistics to develop and to test hypotheses on the relationships of variables to gain insight into the preferred content and format of an assessment instrument from the viewpoint of multiple constituents and stakeholders, if the university so desires. Additionally, the AIDA model includes provisions for the use of inferential statistics to support explanatory research. On the other hand, if the university does not want to invest the time and effort to use inferential statistics for explanatory research purposes, the AIDA model provides a comprehensive approach for exploratory research using qualitative and/or other quantitative methods.

Along with including an option for using hypotheses testing, the AIDA model incorporated a review theory and concepts from which relevant dimensions could be derived from the major constructs using a deductive, top-down approach. In turn, these dimensions could be broken down further to develop appropriate questions to include in interview guides and items to include in the surveys. This top-down approach also served as an effective tool to develop the measurement scale for the surveys. The linkage between the bottom-up and top-down approaches is another positive aspect of the AIDA model because it provides an effective means to build more consistent storylines and account for any inconsistencies in the data.

While there are several positive aspects to the AIDA model, there are several negative aspects or challenges associated with the application of this model. Applying the AIDA model is a relatively complex and time-consuming approach, particularly considering it is only a sub-process in the development of a comprehensive assessment system. Referring to Chapter 3 of this document, prominent authors on performance assessment (Bernardin, 1986; Bernardin & Beatty, 1984; Latham & Wexler, 1994; Mohrman, Resnick-West, and Lawler, 1989) suggest a multi-step process for developing an assessment system. Ingram and Weary (2000), Nason (1997), and Schwartz (1998, 2001) provide additional considerations for developing a useful president assessment system. Even though this study took into account the writings of these authors in the development of the AIDA model, a university should apply the model in context with the other steps in the development of an assessment system for the purposes of compatibility and continuity. Applying the AIDA model as an independent approach could significantly reduce its credibility and utility.

One of the challenges researchers will face in applying the entire AIDA model is managing, assimilating, retrieving, analyzing, and presenting the large volume of data that it generates. For example in this study, the Pearson product-moment correlation coefficient matrix for the pilot study data was very large (111x111), making it very difficult to interpret given the display limitations of a computer and paper printouts from SPSS[®]. Similarly, there were 368 different responses to the initial individual interviews and focus group interview questions. Researchers must be prepared to devote extra time and effort if they choose to complete all of the steps in the process.

Turning to the finer details of the methodology used in this study in support of the AIDA model, using faculty members for the target population poses its own challenges. Arranging for individual interviews and focus group interviews usually requires multiple follow-ups and results in unanticipated delays in data collection. For this study, the invitations for the pilot survey and final survey were sent out via the all faculty list serves. The response rates to invitations sent out using the list serves were relatively low because many faculty members do not take the time to read mass mailings or the invitation becomes “buried” among the myriad of other e-mails. Considering these challenges, researchers should be prepared for the extra time it may take to collect data from faculty members or identify more efficient ways to accomplish the associated tasks.

One final negative aspect of the AIDA model was conducting the pilot survey in parallel with the initial interviews. Had the initial interviews been conducted before the pilot survey, the data could have been used to improve the quality of the pilot survey. For example, the interviewees could have performed a pre-test of the pilot survey during the interview session and provided immediate feedback on the content and format of the pilot survey. A pre-test of the pilot survey might have resulted in a more focused and shorter survey instrument in which more faculty members would have been more inclined to complete.

Convergent Conclusions on Assessment Instrument Development

The primary areas in which the conclusions of this study converge with literature are as follows:

- 1) the abundance of useful literature on theories and concepts that can be applied to develop the content and format of an assessment instrument, 2) the importance of

having a structured and systematic approach for developing the content and format of an assessment instrument, and 3) the challenges associated with development of an assessment instrument that will produce reliable and valid data.

Reinforcing the finding on the abundance of literature, Bass (2008) notes there were over 55,000 publications on leadership in the Online Computer Library Center (OCLC) in 1999 and over 18,000 books were on sale through Amazon.com™ by 2005. In 2005, “Google™ Scholar listed 95,500 publications on leadership, 16,800 books on leadership, and 386,000 citations related to leadership” (Bass, 2008, p. 6). The volume of information on evaluation and assessment of individuals and organizations is also expansive. Searching Google™ in 2009, there were over 2200 publications on the subjects of performance appraisal, individual performance assessment, organizational evaluation, performance management, and performance measurement. Amazon.com™ listed over 94,000 books that address some aspect of these subjects. While these subjects overlap and many of the books are listed in multiple categories, the volume of publications is still very large. Turning to the university presidents, Google™ Scholar lists 41 publications on university presidents and Amazon.com™ lists over 46,000 books containing information on the subject. However, the amount of literature on university president appraisal or assessment publications is relatively small with Google™ Scholar listing only 386 publications and Amazon.com™ listing only 117 books. Considering the large volume of academic publications on leadership, Padilla (2005) emphasizes that “the scope of what has been written about leadership presents a challenge of distillation and synthesis if one is to avoid a representation that is too simplistic, too trivial, or too much of a condensation” (p. 40). The issue challenge of distilling and synthesizing the

information applies to individual performance assessment as well. On the other hand, since there is a limited amount of authoritative literature available on developing performance assessment instruments for university presidents, this study provides results that can broaden the information base on this subject.

Many authors (Bernardin, 1986; Fine, 1986; Latham & Wexley, 1994; Mohrman, Resnick-West, & Lawler, 1989) emphasize the importance of having a systematic approach to developing performance assessment systems, more recently being referred to as performance management systems (Armstrong, 2009). One of the more commonly cited approaches (Armstrong, 2009; Latham & Wexley, 1994; Sokol & Oresick, 1986) for adding structure to assessment instrument development is critical incident technique (CIT) (Flanagan, 1954). As a form of job competency analysis, CIT is very useful in identifying behaviors that can serve as assessment criteria (Armstrong, 2009). CIT came about in the 1950s because of concerns over the use of merit ratings (trait assessments) for performance appraisals (Armstrong, 2009) as they are too subjective and open to prejudicial judgments.

Serving as the primary qualitative method applied in this study, grounded theory (Charmaz, 2006; Strauss & Corbin, 1990, 1998) provided a structured approach for condensing the vast amount of information on theories and concepts associated with university president leadership into a useable form to refine the data collection instruments for this study by using various coding techniques. Content analysis (Klenke, 2008) served as a useful supplement to grounded theory procedures in the coding of comments from the final survey. Applying a bottom-up, constructivist approach to

developing meaning, grounded theory also contributed to the development of hypotheses derived from the this study that could be tested in future research (Charmaz, 2006).

As the main quantitative method used in this study, factor analysis (Field, 2005; Pett, Lackey, & Sullivan) provided a means to perform structural analysis of the phenomenon of university president assessment. Factor analysis was useful in identifying the interrelationships among a large set of variables and then, through data reduction, grouping the factors with common characteristics (Nunnally & Bernstein, 1994). From the very large set of potential assessment criteria for a university president, factor analysis was useful in identifying criteria to include in a president assessment instrument and in controlling the length of the instrument. Furthermore, these criteria were tailored to the feedback from the faculty who many believe should play a key role in the assessment of their presidents (American Association of University Professors, 2006). As a complement to grounded theory, factor analysis proved to be valuable in developing hypotheses pertaining to the relationships between variables (Nunnally, 1978) that corresponded to potential performance assessment criteria to set the stage for follow-up research.

As methodology that has grown in popularity over the last decade (Creswell & Plano-Clark, 2007), mixed-methods research designs offer a systematic approach that approaches a problem from multiple perspectives. One of the major advantages of mixed-methods research is the systematic combination of bottom-up, inductive reasoning and top-down, deductive reasoning to problem solving. Another key advantage is the ability to corroborate the results of the qualitative and quantitative methods in the form of data triangulation for validity purposes. The complementary aspect of mixed-methods

also provides data to fill knowledge gaps that would occur if the research applied only a qualitative or quantitative method. Mason (as cited in Klenke, 2006) identifies several other advantages of mixed-methods research suggesting that it: 1) encourages “outside-the-box” thinking, 2) increases that capacity for theorizing, 3) enhances or extends the logic of qualitative explanations, 4) provides stronger inferences resulting from quantitative analysis, and 5) provides more diverse views (p. 158).

Addressing the challenges of assessing the performance of leaders such as university presidents, Padilla (2005) suggests that researchers have placed too much emphasis in the past on using quantitative methods exclusively to identify leadership descriptors. Given the complexity of the relationships and interactions among the leaders, followers, and organization, Padilla expresses skepticism in using quantitative methods such as multivariate statistics exclusively to “reach into the deeper structures of leadership phenomena” (p. 70). Klenke (2008) agrees that the study of leadership is particularly well suited for qualitative analysis “because of the interdisciplinary nature of the field which has to be more open about paradigmatic assumptions, methodological preferences, and ideological commitments than many single disciplines” (p. 4). Klenke goes on to say quantitative methods have come under scrutiny by the leadership research community because of dissatisfaction with: 1) the complexity of multivariate models, 2) distribution restrictions (i.e., requirement for multivariate normality), 3) large sample size requirements, and 4) difficulty understanding and interpreting results (Cepeda and Martin, 2005, p. 851). Considering the limitations of quantitative methods, Padilla emphasizes that qualitative methods are well suited for the exploration of multiple levels of relationships and interactions among leaders, followers, and organizations. Klenke

reinforces Padilla's perspective by stating that "qualitative leadership studies, when conducted with the same degree of rigor and concern for quality, have several distinct advantages over quantitative approaches by offering more opportunities to study the phenomena in significant depth" (p. 5).

Another challenge associated with the AIDA model was the additional time and effort required to incorporate a mixed-methods design. Creswell and Plano-Clark (2007) point out that "mixed-methods studies are complex and may require extensive time, resources, and effort on the part of the researcher" (p. 281). They suggest that the following questions should be addressed before making a decision to undertake a mixed-methods design: 1) is there sufficient time to collect and analyze two types of data, 2) are there sufficient resources from which to collect and analyze quantitative and qualitative data, and 3) are the skills and personnel necessary to complete this study available?" (p. 181). Creswell and Plano-Clark also suggest that researchers work in teams because of the increasing demands of mixed-methods designs and the advantage of bringing together individuals with diverse methodological and content expertise. The importance of using a team approach to developing performance assessment systems is reinforced by several authors (Bernardin, 1986; Bernardin & Beatty, 1984; Grote, 2002; Mohrman, Resnick-West, & Lawler, 1989).

The incorporation of mixed-methods into the AIDA model brings about the added requirement of checking the validity of qualitative and quantitative data. While the processes for confirming the validity of these two types of data can be relatively complex and time-consuming in their own right, "the very act of combining qualitative and quantitative approaches raises additional validity issues" (Creswell & Plano-Clark, 2007,

p. 145). The implication is that the process of confirming the quality of two types of data is even more complex and time consuming, and this proved to be the case for this study. The American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (1999) stress that validity is “the most fundamental consideration in developing and evaluating tests” (p. 9). Furthermore, they stress that “a sound validity argument integrates various strands of evidence into a coherent account of the degree to which the existing evidence and theory support the intended interpretation of test scores for specific uses” (p. 17).

Since performance assessment is considered a test to which AERA, APA, and NCME standards should be applied (Nathan & Cascio, 1986), there is an implication that researchers must consider data quality checking from beginning to end of the instrument development process. The AIDA model incorporated an integrated and on-going process of checking the trustworthiness of qualitative data and validity of quantitative data. Even though this approach required extra time and effort, it provided insights into validity earlier in the process so adjustments could be made to increase the quality of the products from the study.

Since leadership performance can be influenced by number situational factors (Armstrong, 2008; Finkelstein, Hambrick, & Cannella, 2009; Latham & Wexley, 1994; Padilla, 2005) it is important that these factors be taken into account to ensure an assessment is fair and equitable. The AIDA model employed during this studied served as a useful means to identify more important assessment criteria and external factors that could affect performance ratings from the perspective of faculty members at UNM and NMSU. The AIDA model could also be applied by other universities to identify

performance criteria and external factors that take into consideration the perspectives of key constituents/stakeholders and relevant conditions that could affect the performance of their presidents. Nason (1997) emphasizes the importance of university president assessment taking into account the institutional context that includes the unique characteristics of the university and its governance structure. He believes that “to assess the president’s actions, or failures to act, by themselves is grossly unfair” (p. 32).

Divergent conclusions on assessment instrument development. The main shortcoming of the AIDA model is that it focuses on developing the content and format of an assessment instrument for a university president rather than the design of a performance assessment system overall. Mohrman, Resnick-West, and Lawler (1989) and others offer appraisal system design models that include steps before and after those proposed in the AIDA model that concentrate on the design of the assessment instrument itself. For example, Mohrman, Resnick-West and Lawler suggest that assessment system designers take the following actions before designing the procedures and tools for conducting the assessment: 1) identify key players in the design (e.g., top management, human resources professional, lawyers, and system users), 2) decide on organizational structure that will guide the design (e.g., involvement of consultants, human resources professionals, and/or task force, 3) assess the current organizational situation (e.g., existing assessment systems, what is going well, what is causing problems, organizational structure, connections to compensation system, legal considerations, and organizational climate that includes trust, support, and amount of openness), and 4) purpose and objectives of the assessment system (e.g., basis for compensation and awards, performance improvement, documentation of personnel decisions, performance feedback,

and/or individual development or training). After designing and implementing the system, these authors recommend that the system undergoes a preliminary evaluation and ongoing monitoring to ensure its effectiveness.

Contribution of conclusions to literature on assessment instrument development.

The AIDA model fills a gap that exists in literature on the detailed steps for developing an assessment instrument for a university president. The proper use of this model should provide useful evidence that contributes to the reliability and validity of scores obtained from the application of an assessment instrument.

Assessment Instrument Content and Format Conclusions

The paragraphs that follow in this section address Research Question 2: *What is the preferred content and format for a president performance assessment instrument?*

The answer to this research question is derived from the preliminary, intermediate, and final analysis of qualitative and quantitative data collected during this study. The interviews, pilot survey, and final survey were the data sources for these analyses.

Conclusions on Assessment Instrument Content and Format

This study identified 222 candidate criteria that could serve as content for a university performance assessment instrument. After reviewing 587 publications dating from 1990 to 1989, Rost (1991) found 221 definitions of leadership. Undoubtedly, the number of definitions is much higher today given the proliferation of concepts on leadership (Bass, 2008) over the last two decades.

Given the large number of definitions, blurred distinctions, and overlapping meanings of associated concepts such as management and social influence, one of the biggest challenges of this study was selecting the items to include on the survey there are

hundreds of potential leadership descriptors and performance assessment criteria. Various authors, leadership theorists, and researchers (Armstrong, 2008; Bass, 2008; Grote, 2002; Ingram & Weary, 2000; Padilla, 2005; Nason, 1997; Trachtenberg, 2008) identify a wide range of traits, behaviors, and performance outcomes that could potentially serve as leadership descriptors to include as items in a performance assessment instrument for a university president.

Some authors (Armstrong, 2008; Bass, 2008, Padilla, 2005) point out potential validity issues associated with using traits as leadership criteria because they do not take into account the different situations and they involve more subjective judgments and prejudices. However, regarding the issue of traits, Nason (1997) insists, “any assessment of [university] presidential performance must take some account of character, personality, and style of the president” (p. 38). This study reinforces Nason’s conclusion that traits are important assessment criteria.

This study also supports the conclusion that behaviors and performance outcomes are important assessment criteria. With few exceptions, interview and survey participants perceived that performance outcomes were of lesser importance than traits and behaviors. This finding is consistent with authors who conclude there is a weak relationship between individual performance and organizational performance (Coens & Jenkins, 2002; Illgen & Favero, 1985; Yudof & Busch-Vishniac, 1996).

With respect to external factors, this study supports the conclusion that external factors can have a significant impact on performance ratings (Armstrong, 2009; Bass, 2008; Duderstadt & Womack, 2003; Finkelstein, Hambrick, & Cannella, 2009; Grote, 2002; Illgen & Favero, 1985; Latham & Wexley, 1994; Rhodes, 2001). The literature

review, interviews, pilot survey, and final survey identified 35 external factors that could have negative impact on performance ratings. Since the final survey contained only seven items related to external factors, the importance of other factors could only be determined based on the pilot survey results or the relative frequency with which these items were mentioned in various source documents and during the interviews. The conclusion derived from this study is that several external factors should be taken into account when performing a fair and equitable assessment of university president performance.

The results of this study leads to the conclusion that president assessment should be a formal process with multiple raters providing 360-degree feedback. Study participants believed that members of the board of regents, faculty, students, and alumni should play a role in president assessment. Results also show that assessment should be conducted on a regular basis and involve the use of an instrument with qualitative and quantitative assessment criteria. Since the final survey contained only seven items related to assessment format, the importance of other factors was determined using the results of the pilot survey or the relative frequency with which these items were identified during the literature review and the interviews.

In addition to providing structure for performance assessment, various authors identify additional advantages of formal assessment pointing to the need for an effective process for developing an assessment instrument. Nason (1997) believes that formal assessment can focus attention on the governance structure of the institution by taking into account the attitudes, prerogatives, and behaviors of applicable groups such as the faculty, students, alumni, and legislators. Nason also suggests that formal assessment

provides: 1) an event that triggers the review and resetting of institutional goals or objectives, 2) a rational, orderly, and systematic way of assessing performance using clearly articulated criteria, 3) a way of strengthening and lengthening president terms in the office if properly done, 4) a means for strengthening the board's position by making them more familiar with terms, conditions, and expectations under which the president was appointed and constraints that can affect performance, and 5) accountability for the president's actions, "particularly in public colleges and universities where taxpayers are sometimes suspicious and politicians have occasionally tried to interfere" (pp. 17-19).

Convergent conclusions on assessment instrument content and format. Table 17 through Table 20 provide considerable evidence of the convergence of findings between the final survey, pilot survey, interviews, and literature review on assessment instrument content and format. Additionally, factor analysis indicated there were eight distinct factors (i.e., strategic leadership, consideration, continuous improvement, university mission support, interpersonal competence, stewardship, academic quality, and responsibility) related to potential assessment criteria. With the exception of academic quality and responsibility, these factors tended to be highly correlated indicating that many of the items were essentially dimensions of the same construct. This finding is consistent with authors who conclude that leadership constructs cannot be fully resolved with quantitative methods due to the large number of variables and interaction effects (Illgen & Favero, 1985; Klenke, 2008; Padilla, 2005).

Table 21 consolidates the information in Table 17 and Table 18 into the top 25 assessment instrument content items based on the final survey scores for trait, behavior, and performance outcome criteria. Table 21 also includes the constructs under which

each of 25 items fall based on the preliminary theoretical/conceptual model derived from the literature review and the revised theoretical/conceptual model derived from factor analysis of final survey data. Comparing the naming convention for each item reinforces the finding that leadership constructs can be ambiguous. Observing the large correlation of six of the eight factor scales provides additional evidence of the lack of distinction among the constructs. Since the primary sources used to develop the content of the final survey related to senior leaders in an organization such as a university president, the problem with discriminant validity comes as no surprise.

Table 21. Top 25 Assessment Instrument Content Items

Item	Original Theoretical/ Conceptual Construct	Revised Theoretical/ Conceptual Construct
1. Promotes institutional interests	Followership	Responsibility
2. Makes responsible decisions regarding resource allocation	Management	Stewardship
3. Demonstrates integrity	Leadership	Consideration
4. Displays job competence	Leadership	University Mission Support
5. Makes judicious decisions on administrator selection	Management	Responsibility
6. Secures adequate resources to support the university mission	Performance Outcome	University Mission Support
7. Maintains awareness of the institutional climate	Organization	Stewardship
8. Makes informed decisions based on best available information and research	Learning	Stewardship
9. Maintains effective control of the budget	Organization	Strategic Leadership
10. Recruits high caliber personnel to fill positions in the university	Management	University Mission Support
11. Increases faculty quality	Performance Outcome	Academic Quality
12. Encourages open sharing of information with others	Learning	Consideration
13. Advocates for the university to strengthen higher education	Organization	University Mission Support
14. Promotes positive change	Followership	Strategic Leadership
15. Develops realistic plans to implement strategy	Management	Strategic Leadership
16. Displays professional courtesy	Followership	Consideration
17. Establishes realistic goals and objectives	Organization	Strategic Leadership
18. Builds trusting relationships with constituents/stakeholders	Leadership	Interpersonal Competence
19. Adapts to changes that impact the university	Learning	Strategic Leadership
20. Provides clear direction and expectations to staff	Management	Strategic Leadership
21. Promotes initiatives that contribute to education quality in New Mexico	Organization	Continuous Improvement
22. Promotes negotiation to resolve conflicts	Followership	Strategic Leadership
23. Clears obstacles to progress	Leadership	Strategic Leadership
24. Provides thoughtful responses to constituents/stakeholders	Followership	Strategic Leadership
25. Supports institutional leaders	Followership	Strategic Leadership

Despite the large number of potential assessment criteria and the ambiguity of the constructs related to the central theme of strategic leadership, the qualitative and quantitative analysis for this study captured the criteria that the faculty participants perceived were most important. If an assessment instrument developer need to reduce the number of items from the list of 25 in Table 21, he or she could refer to the eight-factor solution computed for this study and eliminate items within each factor since they are measuring the same construct. The instrument developer could also refer the factor scales and get a sense of what factors that had higher mean scores and choose more items from the higher scoring factors. The developer could also examine the standard deviations of the items within each factor and eliminate those items that had larger variances since there were greater differences in opinion of the relative importance of these items.

Divergent conclusions on assessment instrument content and format. During the grounded theory axial and theoretical coding process for the final survey data related to performance assessment criteria, it was apparent there were additional factors/categories associated with the central theme of strategic leadership such as communication (Blanchard & Associates, 2007; Drucker, 2008), self-mastery (Goleman, 1998), courage (Treasurer, 2008), and legitimization (Bass, 2008). Furthermore, the literature review identified several additional constructs of interest such as connection, creativity, ethics, humility, initiative, innovativeness, optimism, persuasion, philosophy and problem solving, to name a few (Armstrong, 2009; Bass, 2008; Birnbaum, 1992a; Fisher, Tuck, & Wheeler, 1988; Greenleaf, 2002; Gruder, 2008; Keohane, 2006; Trachtenberg, 2008). The findings from this study illustrate that many of the relevant factors pertaining to

assessment of senior leaders were not included in the study. The fact that the rotated eight-factor solution from PCA accounted for only 56.4% of the variance is further evidence that there are other constructs related to the central theme of strategic leadership, although Pett, Lackey, and Sullivan (2003) classify 50% as a “sizable” (p. 102) amount of variance extracted by the components.

Addressing opposition to formal assessment, Armstrong (2008) provides an excellent summary of the academics and commentators who are critical of performance management systems that include individual performance appraisals or performance assessments. Some authors such as Coens & Jenkins (2002) suggest that coaching is a more effective approach for improving the performance of individuals in the workplace. However, when one examines many of the complaints with performance assessment, there appear to be shortcomings that can be overcome if appropriate steps are taken in the design and execution of the assessment system (Armstrong, 2009). In fact, Armstrong suggests that techniques such as coaching and day-to-day feedback to employees are important aspects of an effective performance management system, but they are not a substitute for more formal performance reviews that include appraisals.

Looking specifically at assessment for university presidents, Ingram and Weary (2000), Kauffman, (1978), and Nason (1997) raise issues with formal assessments systems and their accompanying assessment instruments. Ingram and Weary believe using performance surveys in which respondents provide ratings are “inappropriate on several grounds” (p. 19). They say that “rating scales with numerical values for certain leadership characteristics widely miss the mark of connecting many, theoretically discrete, observable behaviors to what *the creators* of such devices believe constitute

good leadership in the aggregate” (p. 19). They go on to say that ratings have many weaknesses with regard to reliability and validity; oversimplify human qualities, complex behaviors, and interactions; and demean and trivialize the academic presidency. Instead, they suggest that well-constructed survey forms can be helpful if they “solicit opinions and insights” (p. 19) through a series of largely open-ended questions to elicit comments and examples.

Kauffman expresses concern that formalizing university president assessment can detract and distract from what “ought” to be done in order to meet the evaluation systems’ expectations rather than the true needs of the institution. While Nason describes several advantages of formal assessment system, he points out some of the perceived disadvantages: 1) it allegedly undermines the authority and status of the office, 2) it politicizes the president’s role, 3) it shortens the president’s term of office because he or she does not want to be subjected to public evaluation, 4) it is an evasion of the board’s responsibility to monitor the health of the institution and the contribution of the president, and 5) it offers a stage for confrontation between all constituencies of the institution.

Contribution of conclusions to literature on assessment instrument content and format. This study identifies candidate assessment criteria for assessment of a university president based on feedback from a key university constituency – the faculty. While literature offers voluminous information on potential assessment criteria, it provides little information of the relative importance or utility of various criteria. Other than the study by Michael, Schwartz, and Balraj (2001) that examined president success indicators, no studies were found that included quantitative assessment of potential president

assessment criteria. Furthermore, this 2001 study only addressed eight items related to president performance.

This study produced an extensive list of external factors that could have an influence on president performance ratings. Due to the scope of this study, the relative importance of all 35 external factors identified through the literature, interviews, and surveys could not be determined accurately. Nevertheless, the perceived importance of external factors such as political pressures, conflicting priorities with those in shared governance, and rater knowledge of president roles, responsibilities, and actions warrant consideration in the assessment process.

From the standpoint of assessment format, this study offers additional evidence of the preference for formal, 360-degree assessment that involves multiple constituencies and stakeholder groups, rather than a centralized assessment conducted only by members of the board of regents. This study reinforces the opinion that university president assessment should include an instrument that contains qualitative and quantitative items and the instrument should be administered on a regular basis. Similar to the issue with external factors, the relative importance of the 46 characteristics of an effective assessment identified during the literature review, interviews, pilot survey, and final survey could not be accurately assessed.

Implications of the Study

Theoretical Implications

The results of this study imply that the central theme of university president performance is strategic leadership in which there are a wide variety of related constructs. Among these related constructs are consideration, continuous improvement, university

mission support, interpersonal competence, stewardship, academic quality, and responsibility as determined through exploratory factor analysis. Grounded theory analysis identified five additional constructs that could potentially relate to strategic leadership: communication, self-mastery, courage, legitimization, and intelligence. Further analysis would be required to investigate the tenability of these constructs. The preliminary theoretical/conceptual model for this study (Figure 1) served as a useful framework for the study; however, the relationship between the variables in the model were tested and confirmed due to the scope of this study. Notwithstanding, the preliminary model provided a structure for identifying pertinent dimensions of the various constructs that led to the selection of items to include in the interview guides, pilot survey, and final survey.

Research Implications

Due to the complexity of the domain of interest, mixed-methods appears to be a viable research option for exploring and confirming the constructs associated with strategic leadership and the multifaceted considerations related to university president assessment. In this study, qualitative research set the stage for quantitative research by identifying candidate constructs and dimensions related to the content of an assessment instrument (i.e., candidate assessment criteria). In turn, the items included in the interview guides and surveys were derived from the dimensions of each construct. Quantitative methods were then applied to survey data to determine if the preliminary constructs and dimensions held up. Since the quantitative analysis identified new construct names with new dimensional alignments, follow-up qualitative analysis was conducted for validation purposes. The integration of reliability and validity analysis

throughout the study served as a useful mechanism for linking the results of quantitative and qualitative analysis.

Qualitative and quantitative analysis was also linked to gain insight into the relative importance of external factors that could influence president performance ratings and the preferred format for president assessment. The items included in the interview guides, pilot survey, and final survey that pertained to external factors and assessment formats were derived from a review of literature and follow-up interviews. While the quantitative data from the final survey provided an indication of the relative importance of seven external factors and seven assessment format characteristics, the qualitative data from the literature review, interviews, and surveys expanded the findings beyond the quantitative items included in the final survey.

Applied Implications

Developing a reliable and valid assessment instrument for a university president is not a “trivial pursuit.” Assessment sponsors should be prepared to make a moderate investment of resources to complete the instrument development process. Once the instrument has been developed and employed, a modest investment will be required to evaluate and update the assessment process to ensure it meets the needs of the university. Since assessment instrument development is only one aspect of an assessment system, it will be important for the assessment sponsors to identify the requirements and to provide the resources to develop and deploy policies and procedures that address the entire assessment system. Furthermore, successful implementation of an effective president assessment system will require buy-in from the board of regents, president, and key constituent/stakeholder groups. The probability of successful implementation of a

president assessment system is related to the level of commitment, investment, and understanding of the assessment process and instrument and the dissemination and utilization of the assessment results.

Limitations of the Study

Several limitations affected the quality of the study. This section addresses the methodology, reliability, and validity limitations. Additionally, this section identifies actions that should be taken in future studies to overcome these limitations.

Methodology

The primary methodology limitations of this study and actions to overcome these limitations are as follows:

- Target population consisted only of UNM and NMSU faculty members. Data collection should be extended to other university constituents/stakeholders.
- Cross-sectional survey limited the scope of the study to a single point in time. Surveys should be conducted at regular intervals to collect data that can be used to update the performance assessment process and the performance assessment instrument.
- Self-administered web-based surveys did not include mechanisms to ensure the participants were valid members of the target population. Having access to the e-mail addresses of the participants would allow the web-based survey administrator to password protect the survey increasing the likelihood that it is accessible only to the target population. Paper surveys could also be mailed to desired participants to reduce the chance of participation by individuals outside the target population.

Reliability and Validity

The primary reliability and validity limitations of this study and actions to overcome these limitations are as follows:

- Reliability of some of the factor scales were relatively low. Additional items should be added to follow-up surveys for the factor scales with less than five items in an attempt to increase Cronbach's alpha.
- Low response rate on the final survey significantly reduced the internal validity of the study results. Researchers should avoid using list serves to send out survey invitations. E-mail or postal addresses should be obtained to send out survey invitations and/or copies of the survey.
- The lack of random sampling of the target population significantly reduced external validity of the study results. Random sampling should be used to reduce threats to external validity.
- Quantitative data was based on perceptions rather than actual performance. Collecting data on actual president performance and actual university performance would provide opportunity to obtain additional validity evidence.
- Many of the items included in the final survey were perceived to be in the range of very important to critically important. This led to negatively skewed distributions for many variables. Since the lack of multivariate normality can have a negative impact on confirmatory factor analysis and other parametric statistics (Field, 2005; Pett, Lackey, & Sullivan, 2003), consideration should be given to using distributional transformations or non-parametric statistical tests (Stevens, 2002) in future research.

Recommendations

The AIDA model illustrated in Figure 11, and described in detail in previous sections of this document, provides a robust process for developing a performance assessment instrument for a university president. Universities that do not have an assessment system that meets the needs of president constituents and other stakeholders should consider using the AIDA model to develop an assessment instrument that produces reliable and valid results. Findings from the practical application of the AIDA model by a university would provide valuable evidence of its utility.

Compared to the AIDA model used for this study (see Figure 4), the proposed model in Figure 11 includes a step for conducting interviews prior to the pilot survey and after the final survey. One of the purposes of conducting interviews prior to the pilot survey is to obtain feedback on a draft of the survey instrument. The main objective of conducting interviews after the final survey is to obtain feedback on the emerging findings and results of the study as part of the reliability and validity assessment process. Additional recommended changes to the original AIDA model include: 1) making the analysis techniques more generic allowing for different types of qualitative and quantitative analysis methods based on researcher preferences and 2) simplifying the descriptions of the outputs from the various analysis techniques. One final consideration with regard to the AIDA model is that its focus is limited to developing the content and format of the assessment instrument. As such, assessment system developers should consider integrating the AIDA model with additional steps in performance assessment system development proposed by authors such as Bernardin and Beatty (1984), Bernardin

(as cited in Berk, 1986), and Mohrman, Resnick-West, and Lawler (1989) to ensure they apply the AIDA model in the appropriate context.

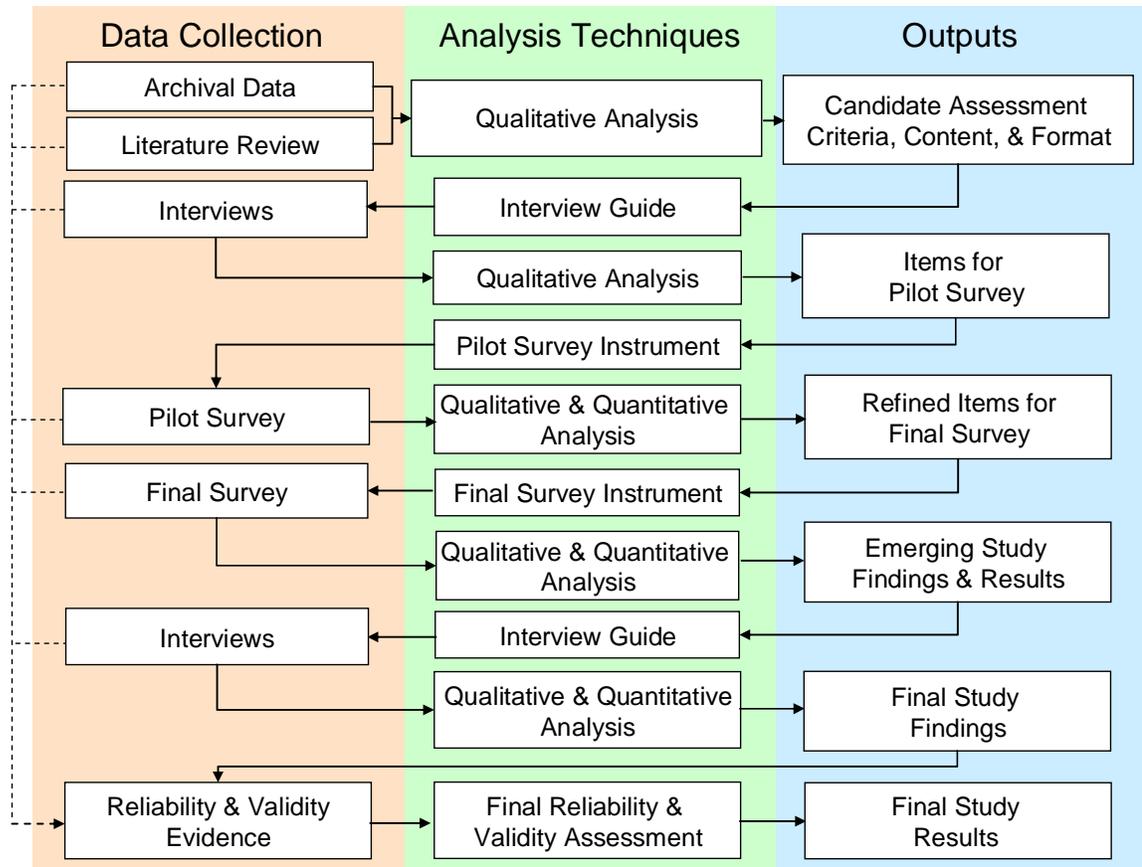


Figure 11. Proposed AIDA Model

Additional research should be conducted to identify relevant constructs and dimensions associated with university president leadership and to confirm the relationships among the pertinent variables. It would be useful to verify the stability of the factors extracted in this study by surveying additional target populations such as university administrators, staff, students, and alumni and by increasing the number of items for the factors that had fewer than five moderate loadings. Investigating the relationships among variables in detail and testing hypotheses related to these

relationships in future research should incorporate confirmatory factor analysis (CFA) techniques such as structural equation modeling [SEM] (Byrne, 2010; Pett, Lackey, & Sullivan, 2003; Nunnally, 1978).

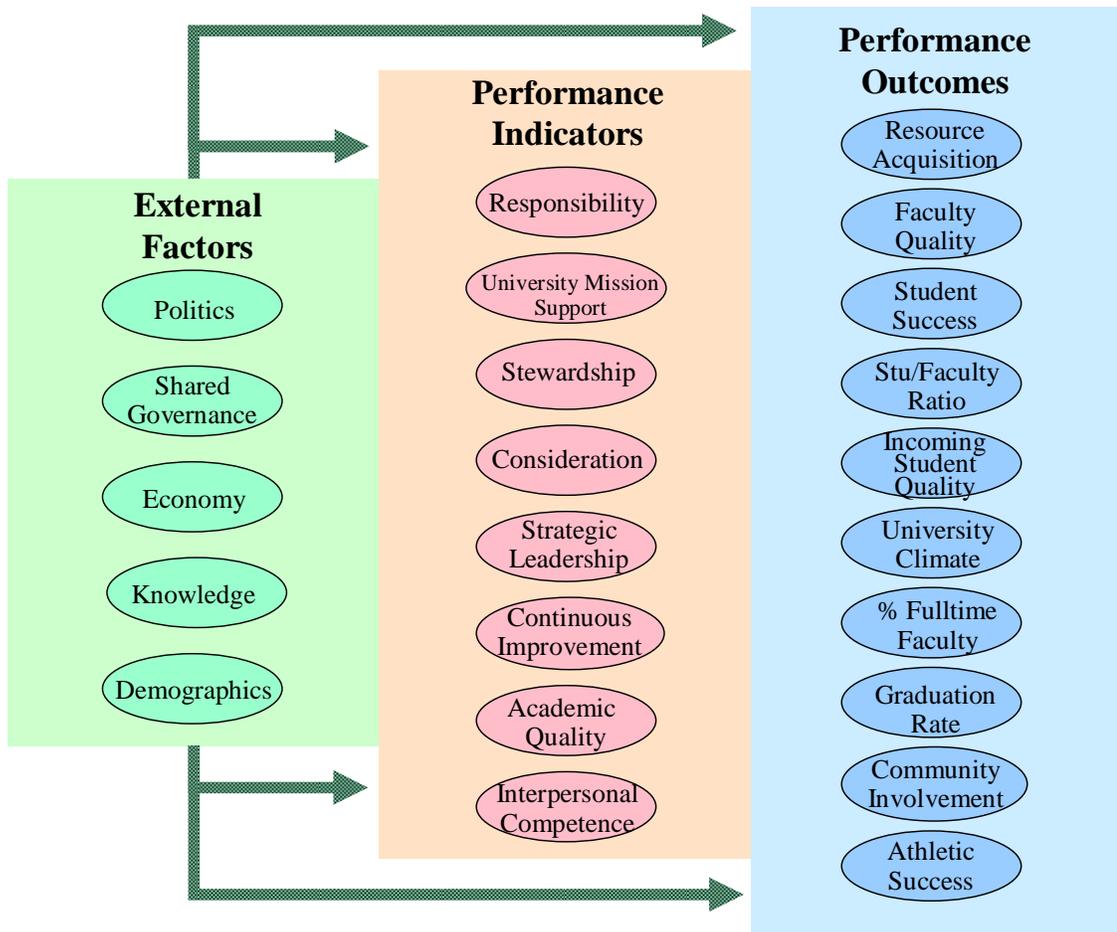


Figure 12. Proposed Theoretical and Conceptual Framework

The proposed theoretical and conceptual framework in Figure 12 should be tested to confirm the relationships among the key variable identified in this study. An example hypothesis that could be used to test the strength of the relationships among variables is as follows (reference): “(University constituent/stakeholder group) will perceive that president performance indicator (X) will have a stronger relationship with university

performance outcome (Y) compared to president performance indicators (T, U, V, and W).”

It would also be useful to develop and maintain a database of actual president performance ratings based on various criteria related to traits and behaviors and compare these ratings to university-level performance outcomes. Comparing performance ratings based on president traits and behaviors to university-level outcomes could provide insight into the impact a president has on university success. This information could then be used to determine if it is appropriate to include university performance outcomes in a president assessment instrument.

Using a multivariate analysis techniques such as a doubly multivariate repeated-measures design (Stevens, 2002), researchers could measure president performance and university performance on several variables over time and could assess performance trends. Example hypotheses that could be used to test the strength of the relationship between president performance and university performance and the moderating effects of external factors are as follows (reference):

- Among university presidents, the performance rating for criterion (X_4) will have a stronger relationship with university performance outcome (Y_1 , Y_2 , and Y_3) compared to president performance rating criteria (X_1 , X_2 , and X_3)
- Among university presidents, the effects of performance outcome (Y_1 , Y_2 , and Y_3), if affected by performance indicator (X_1 , X_2 , and X_3), are moderated by (Z_1 , Z_2 , and Z_3)

Testing these two hypotheses could pose a significant challenge since it would require data collection at multiple universities that use the same president and university

performance criteria and standard scale for measuring variables associated with external factors. The feasibility of having multiple universities adopt a president assessment instrument that includes at least some of the same measures for president performance, university performance, and external factor would have to be determined prior to testing these hypotheses. For example, the amount of resources and commitment necessary to conduct this type of analysis could be prohibitive. Moreover, the potential large number of variables and interactive effects could make it difficult to discern significant relationships as a basis for selecting the best criteria to include in an assessment instrument. Despite the challenges, additional understanding of the relationships between individual performance and organizational performance could be of great benefit to universities and many other types of organizations.

Summary of Discussion

The AIDA model served as an effective tool to guide the research process for this study addressing Research Question 1: *What approach can UNM and other universities use to develop an effective performance assessment instrument for their presidents?* While the AIDA model was applied to developing the data collection instruments and analysis the data for this study, it has applicability to developing an assessment instrument for a university president. Applying the AIDA model in an actual university setting would help confirm its utility and the resources required for successful implementation.

For this study, application of the AIDA model provided the qualitative and quantitative data necessary to answer Research Question 2: *What is the preferred content and format for a president performance assessment instrument?* Research revealed there

are a large number of criteria that can be used as content for a president performance assessment instrument. This study identified 222 traits, behaviors, and performance outcomes that could serve as assessment criteria based on the literature review, interviews, pilot survey, and final survey. Since it is impractical to include a large number of performance criteria in an assessment instrument, the mean scores of the perceived importance of the various criteria on the parts of the final survey participants was used as the primary means to identify the top candidate criteria. Mean scores from the pilot survey and the frequency at which the criteria were mentioned in literature served as secondary sources of information for determining the top 25 criteria.

Another technique used to reduce the number of candidate criteria was exploratory factor analysis in the form of principal components analysis (PCA). PCA was used to determine a smaller set of factors that could adequately describe the candidate assessment criteria. PCA identified an eight-factor solution: consideration, continuous improvement, university mission support, interpersonal competence, stewardship, academic quality, and responsibility. Grounded theory analysis identified five additional constructs (factors) that could potentially relate to strategic leadership: communication, self-mastery, courage, legitimization, and intelligence. Future research should focus on confirming the relationships among these factors and on exploring the potential for limiting the number of items in an assessment instrument to those deemed most useful by members of the board of regents, presidents, and other university constituents/stakeholders.

The results of this study also provided insight into the perceived importance of external factors that could affect president performance ratings. Study participants

believed that political pressures, conflicting priorities with those in shared governance, and rater knowledge of president roles, responsibilities, and actions should be considered in the assessment of president performance. Similarly, the study revealed the preferred format of an assessment instrument in terms of the: 1) degree formality of the system in which it is administered, 2) purpose, 3) structure (i.e., qualitative and quantitative criteria), 4) participants involved in the assessment, and 5) frequency at which the instrument is administered. Based on the scores for the final survey and interviews, study participants believed that universities should perform 360-degree assessments using instruments with qualitative criteria (open-ended questions) and quantitative criteria (closed-ended questions). Additionally, members of the Board of Regents, faculty, staff, students, and the president should be included as raters for assessments conducted on an annual basis.

While this study contributed to the body of research on strategic leadership and university president assessment, there is an opportunity to explore these topics in more depth through follow-up research. Due to the complexity of these subjects, future research should use a mixed-methods design to confirm the relationships of key variables identified in this study and explore the relationships of other variables related to president performance and university performance. Having a better understanding of these variables and their relationships will enable the development of performance assessment instruments for university presidents that will produce more reliable, valid, and useful results.

APPENDICES

APPENDIX A	DEFINITION OF TERMS	252
APPENDIX B	INDIVIDUAL INTERVIEW MATERIALS.....	262
APPENDIX C	FOCUS GROUP INTERVIEW MATERIALS	268
APPENDIX D	PILOT SURVEY	274
APPENDIX E	FINAL SURVEY	281
APPENDIX F	CONSENT FORMS.....	289
APPENDIX G	DATA COLLECTION FORMS.....	294
APPENDIX H	INSTITUTIONAL REVIEW BOARD (IRB) APPROVALS.....	297
APPENDIX I	INITIAL INTERVIEW FINDINGS	306
APPENDIX J	PILOT SURVEY FINDINGS	321
APPENDIX K	FINAL SURVEY FINDINGS	325
APPENDIX L	FACTOR ANALYSIS FINDINGS	332

APPENDIX A
DEFINITION OF TERMS

Analysis of Variance (ANOVA): A statistical procedure that uses the *F*-ratio to test the overall fit of a linear model. In experimental research, this linear model tends to be defined in terms of group means and the resulting ANOVA is therefore an overall test of whether group means differ (Field, 2005).

Assessment criteria: Specific areas in which the evaluators rate president performance using a Likert scale.

Assessment instrument: The document that evaluators use to rate president performance. This document provides instructions, definitions, rating criteria, rating scale, and space for answers to questions that pertain to president performance and effectiveness.

Axial coding: A technique used in grounded theory that relates categories, subcategories, properties, and dimensions of a category (Charmaz, 2006).

Behavior: The way a person behaves or acts (e.g., conduct or manners) (Agnes & Guralnik, 2001). Respondent behavior is unconscious or involuntary reaction to a stimulus and operant behavior is conscious or voluntary reaction to a stimulus (Skinner, as cited in Driscoll, 2000). Certain types of behaviors are frequently used as performance assessment criteria.

Board of Regents: The body that is responsible for the governance of UNM that includes fiduciary responsibility, establishing goals and policies to guide the university, and overseeing the functioning of the university (UNM, 2004).

Coding: “The analytical processes through which data are fractured, conceptualized, and integrated to form theory (Strauss & Corbin, 1998, p. 3).

Category: An abstract grouping of concepts that are derived from qualitative data collected from research participants (Strauss & Corbin, 1990). Fundraising may be an example of a category that has several subcategories such as fundraising through state legislators, research grants, endowments, business investments, and athletic ticket sales.

Constituents: Those individuals or groups that are involved in shared governance at a university including members of the board of regents (or board of trustees) and members of the faculty including deans and department chairs. Some refer to these individuals or groups as stakeholders. See the definition of stakeholders for other individuals and groups who may have an interest in assessment of the president.

Construct: An abstract or hard-to-observe quality or attribute of a person such as their intelligence or leadership ability (Thorndike, 2005). It is also referred to as a latent variable. A construct is equivalent to a category in qualitative research and to a component or factor in quantitative research.

Constructivism: A bottom-up form of inquiry in which meaning and theories are formed from subjective views of individuals based on their personal histories or social interactions with others (Creswell & Plano-Clark, 2007). Constructivism is normally associated with qualitative methods and involves inductive (bottom-up) reasoning (Creswell & Plano-Clark, 2007).

Data Triangulation: Use of a variety of data sources in a study (Denzin as cited in Tashakkori & Teddlie, 1998) to increase the validity or credibility of the results.

Example data sources include surveys, observations, interviews, artifacts, documents, and records (Creswell, 2007).

Dependent Variable (DV): A variable that is not manipulated by the researcher so its value depends on variables that have been manipulated (i.e., independent variables) (Field, 2005).

Descriptive Statistics: Procedures for summarizing a group of scores or otherwise making them more comprehensible (Aron, A. & Aron, E.N, 2003). Descriptive statistics typically include means, standard deviations, z-scores, frequencies, sample sizes, and correlation coefficients that are displayed in tables and graphics.

Detail: The degree to which the assessment instrument contains both objective and subjective assessment criteria that pertain to presidential traits, behaviors, and performance outcomes.

Dimension: A specific attribute or characteristic of a construct that can be observed and measured (Pett, Lackey, & Sullivan, 2003). In the context of a theoretical or conceptual framework, a dimension is a facet or element of a construct (Thorndike, 2005). An item included in a survey is equivalent to a dimension in that it serves as an empirical indicator of a construct. In grounded theory, a dimension is the location of a property of a category along a continuum (Strauss & Corbin, 1990). For example, under the *category* of leadership skills, the *subcategory* is decision-making, with the *property* being importance, and the *dimension* being a range from not important to critically important.

Domain: A general area of interest. For this research, the domain is assessment of university presidents.

Elaboration: A list of specific values or subcategories of a categorical variable or the range of values for a continuous (interval or ratio) variable (Charters, 1992).

External Model: The relationships among different constructs that are described in the theoretical or conceptual framework. The external model specifies the hypotheses to be tested and is evaluated through the process of reviewing evidence of construct validity (Shepard, 1993; Thorndike, 2005).

Formality: The degree to which a structured process exists for assessment of president performance. A formal assessment has written policies, processes, procedures, and assessment instrument.

Format: The dictionary defines format as “general arrangement or plan” (Agnes & Guralnik, 2001, p. 556). In the context of this study, format pertains to:

1) formality (i.e., a formal assessment includes documented policies, procedures, assessment instrument, and report) 2) purpose of the assessment, 3) structure of the assessment instrument (e.g., closed questions, open questions, objective criteria, subjective criteria), 4) participants in the assessment, and 5) frequency at which the assessment instrument should be applied.

Grounded Theory: A qualitative research approach that uses a systematic set of procedures to develop an inductively derived theory about a phenomenon (Strauss & Corbin, 1990).

Independent Variable (IV): A variable that is manipulated by the researcher so its value does not depend on other variables (Field, 2005).

Inferential Statistics: Procedures used in quantitative analysis for drawing conclusions based on scores collected in a research study (sample scores) but going beyond

them to draw conclusions about a population (Aron, A. & Aron, E.N., 2003).

Inferential statistics use various test statistics such as the *t*-Test, *F*- test, Hotelling's T^2 test, Wilk's Λ , and Sheffé's multiple comparison procedures to support hypothesis testing.

Initial Coding: The first step in grounded theory coding involving the naming of each word, line, or segment of data. (Charmaz, 2006).

Internal Model: Interconnected facets or dimensions of a construct, which include all elements to define a construct such as dimensionality, content, and structure (Shepard, 1993; Thorndike, 2005).

Job Analysis: A method used to identify the knowledge, skills, and abilities that contribute to the quality and quantity of work performance by an individual (Fine, 1986). Job analysis can be used to identify the criteria and standards for performance assessment.

Methodological Triangulation: Use of both qualitative and quantitative methods to study the same phenomena within the same study or in different complementary studies (Denzin as cited in Tashakkori & Teddlie, 1998).

Methodology: "A way of thinking about and studying social reality" (Strauss & Corbin, 1998, p. 3).

Methods: "A set of procedures and techniques for gathering and analyzing data" (Strauss & Corbin, 1998, p. 3).

Mixed Methods: The combination of qualitative and quantitative approaches in research methodology for a single study multiphase study (Tashakkori & Teddlie, 1998).

Mode of Variation: The way in which a quantitative measure varies (i.e., a categorical/nominal variable varies in kind and an interval/ratio/continuous variable varies in degree) (Charters, 1992). For example, gender is a categorical variable and height is an interval variable.

Moderator variable (MoV): A nominal or continuous variable that effects the direction and/or strength of the relation between the IV and DV (Baron & Kenny, 1986). In ANOVA or Multiple Regression, a moderator effect is represented by a significant interaction between the IV and MoV. The research hypothesis for a MoV implies a causal relationship between the IV and DV and changes with the presence of a MoV.

Multivariate Analysis of Variance (MANOVA): A family of statistical tests that extend basic ANOVA to situations in which more than one outcome (dependent) variable has been measured (Field, 2005).

Multiple Regression: A statistical test used to analyze the collective or separate effects of two or more independent variables on a single dependent variable (Pedhazur, 1997). This test overlooks the intercorrelation of the independent variables. Multiple regression can be used to analyze the interaction effects of multiple independent variables only if the variables are categorical (or nominal).

Multivariate Multiple Regression: A statistical test used to analyze the collective or separate effects of two or more independent variables on two or more dependent variables (Pedhazur, 1997).

Performance Outcome: The result of an individual, team, group, or organization performing assigned tasks such as the achievement of predetermined goals and

objectives. Certain types of performance outcomes are frequently used as performance assessment criteria.

Postpositivism: The worldview that knowledge is based on: 1) determinism or cause-and-effect thinking, 2) reductionism, by narrowing and focusing on selected variables, 3) observation and measures of variables, and 4) testing theories that are continually refined (Slife & Williams, 1995). Postpositivism is normally associated with quantitative methods and deductive (top-down) reasoning (Creswell & Plano-Clark, 2007).

Pragmatism: The worldview that focuses on the consequences of research with primary importance placed on the research questions rather than the methods (Creswell & Plano-Clark, 2007). Applying the concept of pluralism, pragmatism involves multiple methods of data collection to investigate problems under study and involves deductive and inductive reasoning (Creswell & Plano-Clark, 2007).

President: The chief executive officer (CEO) who is responsible to the board of regents for the operations and management of the university (UNM, 2004).

Principle Component Analysis (PCA): A statistical technique for identifying underlying similarities between groups or variables (George & Mallory, 2007).

Property: Attribute, characteristic, or subcategory pertaining to a phenomenon (Strauss & Corbin, 1990) such as the level of importance of that phenomenon.

Stakeholders: Those individuals, groups, or organizations that may have an interest in the performance and assessment of the president or the research results. For the purpose and scope of this study, primary stakeholders are those who are primarily responsible for shared governance of the university including members of the

board, the president, and faculty. The primary stakeholders are sometimes referred to as constituents. See Table 1 in Chapter 2 for a complete list of stakeholders.

Task: A specific element of work that may be assigned to an individual.

Trait: A stable internal characteristic or tendency of an individual (Thorndike, 2005).

Certain traits are frequently used as performance assessment criteria.

Theoretical Coding: Techniques in grounded theory in which the relationships between the categories are defined as hypotheses that can be integrated into an overarching theory (Charmaz, 2006; Glaser, as cited in Charmaz, 2006).

Theory: An abstract analytical schema of a process, action, or interaction (Strauss & Corbin as cited in Creswell, 2007). Theory consists of the concepts and statements of relationships between those concepts (Strauss & Corbin, 1990). A conceptual (or theoretical) framework consists of a set of constructs (Thorndike, 2005) for a certain domain of interest.

Theory triangulation: A process that uses different theoretical perspectives to interpret the same data. By applying different theories to make sense of data, it is possible to see how different factors such as experiences, assumptions, and beliefs affect the findings (Russ-Eft & Preskill, 2001).

Worldview: Assumptions a researcher makes about reality, how knowledge is obtained, and the methods of gaining knowledge, which is sometimes referred to as a paradigm (Creswell & Plano-Clark, 2007). For the purposes of this study, the three worldviews of interest are *postpositivism*, *constructivism*, and *pragmatism*.

Table A1. Comparison of Terms for this Study

Theoretical/ Conceptual Framework	Qualitative Analysis Grounded Theory	Quantitative Analysis	Performance Assessment	Assessment Instrument	Practical Example
Construct	Category	Component or Factor	Trait, behavior, or task performance outcome	Categories	Leadership Behavior
Dimension	Subcategory	Variable	Assessment Criteria	Items	Demonstration of personal integrity
Importance	Property	Mode of Variation		Rating Scale	Degree of importance of this criteria
Range of Importance	Dimension	Elaboration	Performance Standard	Rating Scale	Likert scale from 1-5 (not important to critically important)

APPENDIX B
INDIVIDUAL INTERVIEW MATERIALS

(Date)

Dear UNM Faculty Member:

I am requesting your participation in an interview that will focus on performance assessment of university presidents. This interview is in support of a study on the development of the content and format of an assessment instrument that will produce reliable and valid results. The University of New Mexico (UNM) Office of the President, UNM Faculty Senate, and Association of Governing Boards of Universities and Colleges are interested in the results of this study.

The results of this study will provide useful information to universities as they attempt to provide fair and objective assessments of president performance. Specifically, the results will provide insight into criteria that key university stakeholders believe will accurately reflect the president's knowledge, behaviors, traits, and external factors that may affect performance outcomes. The ultimate aim of this study is to provide information that universities can use for professional development, organizational development, and goal achievement.

After completing the face-to-face interview, you will be asked to complete a pilot survey questionnaire that will be administered by StudentVoice™. Please take a few minutes to answer the questions contained in this pilot survey after reading the instructions provided in the survey. I will compile the data from your survey along with that provided by other participants. I assure you that your individual responses will remain confidential, i.e., your name will not be connected with your answers. The results of the pilot survey will be used to refine the final survey instrument.

I would like to thank you in advance for providing information that will be valuable to your university. Please call me at 505-853-1397 or send an e-mail to me at the following address: dlester@unm.edu to make arrangements for the interview that can be held at a time and place of your convenience. The duration of this interview is approximately one-hour not including the pilot survey which will take approximately 15 minutes to complete using the web-based survey.

Sincerely yours,

Dennis Lester
Ph.D. Candidate
University of New Mexico

College of Education
Educational Leadership and Organizational Learning
MSC05 3040
1 University of New Mexico
Albuquerque, NM 87131-0001

Subject: Developing an Effective Instrument for Assessing the Performance of University Presidents (IRB Protocol # 08-607)

To: UNM (or NMSU) Faculty Member

I am requesting your participation in an individual interview that pertains to performance assessment of university presidents. This interview is in support of a study on the development of the content and format of an assessment instrument that will produce reliable and valid results.

The duration of this interview is approximately 1 hour. The primary purpose of this interview is to review the emerging results of the study and to comment on the survey instrument for this study. I assure you that your individual responses will remain confidential, i.e., your name will not be connected with your answers.

The results of this study will provide useful information to universities as they attempt to provide fair and objective assessments of president performance. Specifically, the results will provide insight into criteria that key university stakeholders believe will accurately reflect the president's knowledge, behaviors, traits, assessment methods, and external factors that may affect performance outcomes. The ultimate aim of this study is to provide information that universities can use for professional development, organizational development, and goal achievement.

I would like to thank you for providing information that will be valuable to the university. Please call me at 505-853-1397 or send an e-mail to me at dlester@unm.edu to arrange for the interview.

Sincerely yours,

Dennis Lester
Ph.D. Candidate
University of New Mexico

College of Education
Educational Leadership and Organizational Learning
MSC05 3040
1 University of New Mexico
Albuquerque, NM 87131-0001

INDIVIDUAL INTERVIEW GUIDE			
Date:	Time:	Interviewer Name:	
Number of Participants:		Location:	
Introduction and Instructions			
<ol style="list-style-type: none"> 1. Welcome and Introductions (name, college, position, years at UNM) 2. Purpose of Interview/General Aim of Activity 3. Session Length (1.5 hours) 4. Administrative information (break at 45 minutes, rest rooms, note taking) 5. Handouts (pilot survey) 6. Sequence of Events 7. Rules of Order (questions, responses, etiquette, staying on track, time allocation) 8. Assignment (discuss pilot survey after completion of questions) 9. Academic Freedom 10. Response recording (handwritten notes and tape recording) 11. Confidentiality/Consent 12. Use of information 13. Evaluation results availability 			
Comments and Questions			
<ol style="list-style-type: none"> 1. We will be using what is referred to as Critical Incident Technique (Flanagan, 1954) to examine the behaviors of university presidents with a slight addition in the scope of the discussion to include traits and performance outcomes. 2. This is a study of assessment of university presidents – we believe you are well qualified to tell us about your experiences in dealing with a university president or to tell us about your perceptions of the traits, behaviors, and performance outcomes that contribute to his or her success. 3. What would you say is the primary purpose or aim of presidential assessment? 4. What do you think are the traits of a good university president? 5. Think of a time when a university president has done something that you personally observed or heard or read about that in your opinion was an example of good behavior. Please explain the behavior and please provide the context (what, when, where, who was involved, and relevant factors bearing situation). 6. Think of a time when a university president has done something that you personally observed or heard or read about that in your opinion was an example of poor behavior. Please explain the behavior and please provide the context (what, when, where, who was involved, and relevant factors bearing situation). 7. What do you think are the more relevant areas in which to assess presidential performance outcomes? 8. What are some of the factors that are beyond the presidents’ control that can affect his or her performance? 			

<p>9. Do you think that a formal assessment of the president is of value to the president and the university? (Formal assessment is defined as having written policies, procedures, assessment criteria, an assessment instrument, and performed on a regular basis).</p> <p>10. If you believe assessment is valuable, who do you think should be involved?</p>
<p>Assignment</p> <ul style="list-style-type: none">▪ Complete the pilot survey prior to departing.
<p>Closing Remarks</p> <ol style="list-style-type: none">1. Thank you for participating2. Reminder on confidentiality and data handling3. Timetable for the remainder of the evaluation4. Availability of products from the evaluation

INTERVIEW GUIDE SURVEY AND INTERVIEW FEEDBACK		
Date:	Time:	Interviewer Name:
Number of Participants:		Location:
<p>Introduction and Instructions</p> <ol style="list-style-type: none"> 1. Welcome and introductions 2. Purpose of interview 3. Session length (1 hour for individual interviews and 1.5 hours for focus group interviews) 4. Administrative information (rest rooms, refreshments, note taking) 5. Handouts (survey form and survey results) 6. Sequence of events 7. Rules of order (questions, responses, etiquette, staying on track, time allocation) 8. Academic freedom 9. Response recording (handwritten notes and tape recording) 10. Confidentiality/consent 11. Use of information 12. Evaluation results availability 		
<p>Questions and Discussion</p> <ol style="list-style-type: none"> 1. Question: Have you taken the Developing an Effective Instrument for Assessing the Performance of Public University Presidents survey? If so, what was your general opinion of it? 2. Discussion: Results of the survey and previous interviews. 3. Question: In your opinion, do the results seem to be reliable and valid? If not, why not? 4. Question: Do you have any suggestions on how to improve the format of the instrument for future use? 5. Question: Do you have any suggestions on how to improve the wording of the items in the survey for future use? 6. Question: Do you have any suggestions on what items should be added or deleted in the survey for future use? 7. Discussion: Summary of suggested improvements. 		
<p>Assignment</p> <ul style="list-style-type: none"> ▪ None 		
<p>Closing Remarks</p> <ol style="list-style-type: none"> 5. Thank you for participating 6. Reminder on confidentiality and data handling 7. Timetable for the remainder of the evaluation 8. Availability of products from the evaluation 		

APPENDIX C
FOCUS GROUP INTERVIEW MATERIALS

(Date)

Dear UNM Faculty Member:

I am requesting your participation in a focus group interview that pertains to performance assessment of university presidents. This interview is in support of a study on the development of the content and format of an assessment instrument that will produce reliable and valid results. The University of New Mexico (UNM) Office of the President, UNM Faculty Senate, and Association of Governing Boards of Universities and Colleges are interested in the results of this study.

The results of this study will provide useful information to universities as they attempt to provide fair and objective assessments of president performance. Specifically, the results will provide insight into criteria that key university stakeholders believe will accurately reflect the president's knowledge, behaviors, traits, and external factors that may affect performance outcomes. The ultimate aim of this study is to provide information that universities can use for professional development, organizational development, and goal achievement.

After completing the face-to-face interview, you will be asked to complete a pilot survey questionnaire that will be administered by StudentVoice™. Please take a few minutes to answer the questions contained in this pilot survey after reading the instructions provided in the survey. I will compile the data from your survey along with that provided by other participants. I assure you that your individual responses will remain confidential, i.e., your name will not be connected with your answers. The results of the pilot survey will be used to refine the final survey instrument.

I would like to thank you in advance for providing information that will be valuable to your university. Please call me at 505-853-1397 or send an e-mail to me at the following address: dlester@unm.edu to make arrangements for the interview that can be held at a time and place that will be pre-coordinated with you and the other participants. The duration of this interview is approximately 1 ½ hours not including the pilot survey which will take approximately 15 minutes to complete using the web-based survey. I anticipate that 10-15 tenured faculty members will participate in this focus group interview.

Sincerely yours,

Dennis Lester
Ph.D. Candidate
University of New Mexico

College of Education
Educational Leadership and Organizational Learning
MSC05 3040
1 University of New Mexico
Albuquerque, NM 87131-0001

Subject: Developing an Effective Instrument for Assessing the Performance of University Presidents (IRB Protocol # 08-607)

To: UNM (or NMSU) Faculty Member

I am requesting your participation in a focus group interview that pertains to performance assessment of university presidents. This interview is in support of a study on the development of the content and format of an assessment instrument that will produce reliable and valid results.

The duration of this interview is approximately 1-½ hours. I anticipate that 3-5 tenured faculty members will participate in this focus group interview. The primary purpose of this interview is to review the emerging results of the study and to comment on the survey instrument. I assure you that your individual responses will remain confidential, i.e., your name will not be connected with your answers.

The results of this study will provide useful information to universities as they attempt to provide fair and objective assessments of president performance. Specifically, the results will provide insight into criteria that key university stakeholders believe will accurately reflect the president's knowledge, behaviors, traits, assessment methods, and external factors that may affect performance outcomes. The ultimate aim of this study is to provide information that universities can use for professional development, organizational development, and goal achievement.

I would like to thank you for providing information that will be valuable to the university. Please call me at 505-853-1397 or send an e-mail to me at dlester@unm.edu to arrange for the interview. The time and place will be pre-coordinated with you and the other participants.

Sincerely yours,

Dennis Lester
Ph.D. Candidate
University of New Mexico

College of Education
Educational Leadership and Organizational Learning
MSC05 3040
1 University of New Mexico
Albuquerque, NM 87131-0001

FOCUS GROUP INTERVIEW GUIDE			
Date:	Time:	Interviewer Name:	
Number of Participants:		Location:	
Introduction and Instructions			
<ul style="list-style-type: none"> 13. Welcome and Introductions (name, college, position, years at UNM) 14. Purpose of Interview/General Aim of Activity 15. Session Length (1.5 hours) 16. Administrative information (break at 45 minutes, rest rooms, refreshments, note taking) 17. Handouts (pilot survey) 18. Sequence of Events 19. Rules of Order (questions, responses, etiquette, staying on track, time allocation) 20. Assignment (discuss pilot survey after completion of questions) 21. Academic Freedom 22. Response recording (handwritten notes and tape recording) 23. Confidentiality/Consent 24. Use of information 25. Evaluation results availability 			
Comments and Questions			
<ul style="list-style-type: none"> 8. We will be using what is referred to as Critical Incident Technique (Flanagan, 1954) to examine the behaviors of university presidents with a slight addition in the scope of the discussion to include traits and performance outcomes. 9. This is a study of assessment of university presidents – we believe you are well qualified to tell us about your experiences in dealing with a university president or to tell us about your perceptions of the traits, behaviors, and performance outcomes that contribute to his or her success. 10. What would you say is the primary purpose or aim of presidential assessment? 11. What do you think are the traits of a good university president? 12. Think of a time when a university president has done something that you personally observed or heard or read about that in your opinion was an example of good behavior. Please explain the behavior and please provide the context (what, when, where, who was involved, and relevant factors bearing situation). 13. Think of a time when a university president has done something that you personally observed or heard or read about that in your opinion was an example of poor behavior. Please explain the behavior and please provide the context (what, when, where, who was involved, and relevant factors bearing situation). 14. What do you think are the more relevant areas in which to assess presidential performance outcomes? 15. What are some of the factors that are beyond the presidents’ control that can 			

affect his or her performance?

16. Do you think that a formal assessment of the president is of value to the president and the university? (Formal assessment is defined as having written policies, procedures, assessment criteria, an assessment instrument, and performed on a regular basis).

17. If you believe assessment is valuable, who do you think should be involved?

Assignment

- Complete the pilot survey prior to departing.

Closing Remarks

9. Thank you for participating
10. Reminder on confidentiality and data handling
11. Timetable for the remainder of the evaluation
12. Availability of products from the evaluation

INTERVIEW GUIDE SURVEY AND INTERVIEW FEEDBACK		
Date:	Time:	Interviewer Name:
Number of Participants:		Location:
Introduction and Instructions		
<ul style="list-style-type: none"> 26. Welcome and introductions 27. Purpose of interview 28. Session length (1 hour for individual interviews and 1.5 hours for focus group interviews) 29. Administrative information (rest rooms, refreshments, note taking) 30. Handouts (survey form and survey results) 31. Sequence of events 32. Rules of order (questions, responses, etiquette, staying on track, time allocation) 33. Academic freedom 34. Response recording (handwritten notes and tape recording) 35. Confidentiality/consent 36. Use of information 37. Evaluation results availability 		
Questions and Discussion		
<ul style="list-style-type: none"> 18. Question: Have you taken the Developing an Effective Instrument for Assessing the Performance of Public University Presidents survey? If so, what was your general opinion of it? 19. Discussion: Results of the survey and previous interviews. 20. Question: In your opinion, do the results seem to be reliable and valid? If not, why not? 21. Question: Do you have any suggestions on how to improve the format of the instrument for future use? 22. Question: Do you have any suggestions on how to improve the wording of the items in the survey for future use? 23. Question: Do your have any suggestions on what items should be added or deleted in the survey for future use? 24. Discussion: Summary of suggested improvements. 		
Assignment		
<ul style="list-style-type: none"> ▪ None 		
Closing Remarks		
<ul style="list-style-type: none"> 13. Thank you for participating 14. Reminder on confidentiality and data handling 15. Timetable for the remainder of the evaluation 16. Availability of products from the evaluation 		

APPENDIX D
PILOT SURVEY

PILOT SURVEY COVER LETTER

UNM Letterhead

(Date)

Dear UNM Faculty Member:

I am writing to ask your opinion how to assess the performance of university presidents. I recently sent you a postcard telling you about the survey, and am now providing you with the survey questionnaire for you to complete.

Please take a few minutes to answer the questions contained in this survey after reading the instructions provided in the survey booklet. I will compile the data from your survey along with that provided by other participants. I assure you that your individual responses will remain confidential – that is, your name will not be connected with your answers.

The results of this survey will provide useful information to universities as they attempt to provide fair and objective assessments of president performance. Specifically, the results will provide insight into criteria that key university stakeholders believe will accurately reflect the president's knowledge, behaviors, traits, and external factors that may affect performance outcomes. The ultimate aim of this survey is to provide information that universities can use for professional development, organizational development, and goal achievement.

Please complete your survey as soon as possible. After you complete this survey, please return it in the postage-paid envelope. If you prefer to complete the survey on the Web, please go to <https://www.studentvoice.com/login.asp> and type in your personal User ID and Password printed in the top left corner of this letter. The instructions for completing the Web survey are available once you have logged onto the survey site.

I would like to thank you in advance for providing information that will be valuable to your university. If you have any questions, please call me at 505-853-1397 or send an e-mail to me at the following address: dlester@unm.edu.

Sincerely yours,

Dennis Lester
Ph.D. Candidate
University of New Mexico

College of Education
Educational Leadership and Organizational Learning
MSC05 3040
1 University of New Mexico
Albuquerque, NM 87131-0001

**UNIVERSITY PRESIDENT
ASSESSMENT SURVEY**

PLEASE COMPLETE AND RETURN THIS QUESTIONNAIRE BY **FEBRUARY 15, 2009** USING THE ENVELOPE PROVIDED.

Questions? Call Dennis Lester 505.853.1397 or e-mail to dlester@unm.edu

Instructions:

- Use a No. 2 pencil only.
- Do not use ink, ballpoint pens, or felt tip pens.
- Make solid marks that fill in the circle completely. Correct mark example: ●
- Cleanly erase any marks you wish to change and do not make stray marks on the form.

Please rate the importance of the following factors to the successful performance of a university president. Please mark each item based on its own merit rather than its relative importance compared with other items on the questionnaire. Also, consider the full range of the scale from “not important” to “critically important” in making your choices to increase the likelihood of identifying true differences among these factors during data analysis.

[Survey barcode]	Critically Important	5					
	Very Important	4					
	Important	3					
	Slightly Important	2					
	Not Important	1	1	2	3	4	5
1. The <u>President</u> promotes learning in the following ways:			<input type="radio"/>				
a. Rewards superior performance			<input type="radio"/>				
b. Applies previous knowledge/experience to solve problems			<input type="radio"/>				
c. Encourages open sharing of knowledge among constituents			<input type="radio"/>				
d. Articulates the university story (e.g., vision, mission, values)			<input type="radio"/>				
e. Adapts to changes that impact the university			<input type="radio"/>				
f. Assists individuals in achieving their goals/aspirations			<input type="radio"/>				
g. Promotes harmony among university stakeholders			<input type="radio"/>				
h. Incorporates lesson learned into decision making			<input type="radio"/>				
2. The President displays the following leadership qualities:							
a. Demonstrates assertiveness in resolving issues			<input type="radio"/>				
b. Demonstrates a high degree of personal integrity			<input type="radio"/>				
c. Builds trusting relationships with others			<input type="radio"/>				
d. Displays a high degree of job competence			<input type="radio"/>				
e. Maintains an upbeat attitude			<input type="radio"/>				
f. Suggests novel ways of doing business			<input type="radio"/>				
g. Inspires constituents to follow his or her lead			<input type="radio"/>				
h. Tailors leadership style to the follower expectations			<input type="radio"/>				
i. Promotes self-reflection to transform old ways of thinking			<input type="radio"/>				
j. Displays passion toward his or her work			<input type="radio"/>				
k. Clears obstacles that enable constituents to be successful			<input type="radio"/>				
l. Provides a framework for developing institutional strategy			<input type="radio"/>				
m. Incorporates a holistic systems approach in problem solving			<input type="radio"/>				

Proceed to question 3 on the next page

	Critically Important 5					
	Very Important 4					
	Important 3					
	Slightly Important 2					
	Not Important 1	1	2	3	4	5
3. The <u>President</u> incorporates the following management practices:						
a. Develops realistic plans to implement strategy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Makes logical decisions regarding work assignments for staff	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Provides clear direction for task execution	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Recruits high caliber personnel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Uses appropriate performance indicators to make decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Manages risk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. The <u>President</u> exhibits the qualities of a good follower by:						
a. Serving as an agent for positive change	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Displaying a willingness to reach consensus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Raising controversial issues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Demonstrating professional courtesy to others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Promoting institutional interests rather than self-interests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Providing support to those individuals in leadership roles	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. The <u>President</u> adopts the following organizational practices:						
a. Maintains an effective knowledge management system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Takes appropriate actions to secure necessary resources	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Established realistic goals for the institution	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Maintains effective control over resources	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Performs benchmarking to identify improvements	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Establishes effective teams	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. Maintains good awareness of stakeholder satisfaction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. Promotes initiatives that contribute to society	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. Focuses strategies on increased competitiveness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. Empowers individual by decentralizing authority	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k. Promotes continuous learning for continuous improvement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
l. Emphasizes customer satisfaction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. The <u>University</u> maintains a president assessment system that:						
a. Ties president compensation to president performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Provides ongoing feedback for personal development	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Involves multiple stakeholders in the assessment process	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Includes assessment on a regularly scheduled basis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Documents associated policy/processes/procedures	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Utilizes an assessment instrument with specific criteria	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. Includes objective versus subjective assessment criteria	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Proceed to question 7 on the next page						

	Critically Important	5					
	Very Important	4					
	Important	3					
	Slightly Important	2					
	Not Important	1	1	2	3	4	5
7. Factors that influence presidential performance ratings include:							
a.	Variables in which the president has no control		<input type="radio"/>				
b.	Cultural backgrounds of the university stakeholders		<input type="radio"/>				
c.	Stakeholder desire to maintain autonomy from the staff		<input type="radio"/>				
d.	Scholarly criticism/skepticism within the university		<input type="radio"/>				
e.	Experience of stakeholders involved in the assessment		<input type="radio"/>				
f.	Economic conditions surrounding the university		<input type="radio"/>				
g.	Perspectives of followers on the attributes of a good leader		<input type="radio"/>				
h.	Multiple conflicting goals/priorities of the university		<input type="radio"/>				
i.	Shared governance with university stakeholders		<input type="radio"/>				
j.	Rater errors (e.g., halo effect, leniency, and central tendency)		<input type="radio"/>				
8. <u>President</u> success is based on the following outcomes:							
a.	Increased quality of incoming students		<input type="radio"/>				
b.	Increased success of students		<input type="radio"/>				
c.	Increased research/scholarship		<input type="radio"/>				
d.	Improved faculty credentials		<input type="radio"/>				
e.	Increased faculty productivity		<input type="radio"/>				
f.	Increased fundraising		<input type="radio"/>				
g.	Increased integration of diverse cultures		<input type="radio"/>				
h.	Improved relationships with stakeholders		<input type="radio"/>				
i.	Improved infrastructure (i.e., facilities and equipment)		<input type="radio"/>				
j.	Increase customer satisfaction		<input type="radio"/>				
k.	Improved standing of the university		<input type="radio"/>				
l.	Improved quality of athletic programs		<input type="radio"/>				
m.	Increased involvement in the local community		<input type="radio"/>				
n.	Increased quality of athletic programs		<input type="radio"/>				
o.	Increased accessibility of students to courses		<input type="radio"/>				
p.	Improved quality of academic programs		<input type="radio"/>				
q.	Increased use of advanced information technologies		<input type="radio"/>				
r.	Increased participation in the global community		<input type="radio"/>				
Proceed to question 10 on the next page							

Please provide the following personal information. Fill in circles or blanks as applicable.	
9. What is your age?	<input type="radio"/> 20 – 29 years <input type="radio"/> 30 – 39 years <input type="radio"/> 40 – 49 years <input type="radio"/> 50 – 59 years <input type="radio"/> 65 years or older
10. What is your gender?	<input type="radio"/> Male <input type="radio"/> Female
11. What is your race/ethnicity?	<input type="radio"/> American Indian <input type="radio"/> Black <input type="radio"/> Chinese <input type="radio"/> Hispanic <input type="radio"/> Japanese <input type="radio"/> Korean <input type="radio"/> Vietnamese <input type="radio"/> White <input type="radio"/> Other, please specify:
12. What is your status as a faculty member at UNM?	<input type="radio"/> Full-time faculty <input type="radio"/> Part-time faculty <input type="radio"/> Assistant professor <input type="radio"/> Associate professor <input type="radio"/> Professor <input type="radio"/> Instructional faculty <input type="radio"/> Visiting faculty <input type="radio"/> Clinician educator <input type="radio"/> Temporary faculty <input type="radio"/> Other, please specify:
13. Do you another job in addition to teaching at UNM?	<input type="radio"/> Yes, I have a part-time job <input type="radio"/> Yes, I have a full-time job <input type="radio"/> No
14. Do you have additional comments on indicators of university president success, assessing university presidents, and/or this survey? If so, please write them on the next page.	

Thank you for completing this survey!

Please return your completed survey to: Dennis Lester ♦ P.O. Box 1234 ♦ Albuquerque, NM 871XX-XXXX

APPENDIX E
FINAL SURVEY

Subject: Developing an Effective Instrument for Assessing the Performance of University Presidents (IRB Protocol # 08-607)

To: UNM and NMSU Faculty Members

As a follow up to the pilot survey conducted in March-April 2009 at the University of New Mexico (UNM), you are invited to participate in the final survey being conducted at UNM and New Mexico State University (NMSU). The survey questionnaire will be administered by StudentVoiceTM. The link to the website for completing the survey is <http://studentvoice.com/unm/UnivPresPerfAssmt>.

This survey supports a study that will propose a model for developing an assessment instrument that provides reliable and valid results. It will also identify criteria that university faculty members believe are critical to university president success. The ultimate purpose of this study is to provide information that universities can use for professional development, organizational development, and goal achievement. Its purpose is not to assess the performance of the current presidents of UNM or NMSU or the president of any other university.

The survey should take approximately 15 minutes to complete. Responses will remain confidential and names will not be connected with answers. The consent form is included in the web-based survey.

Thank you for taking the time to participate in this study.

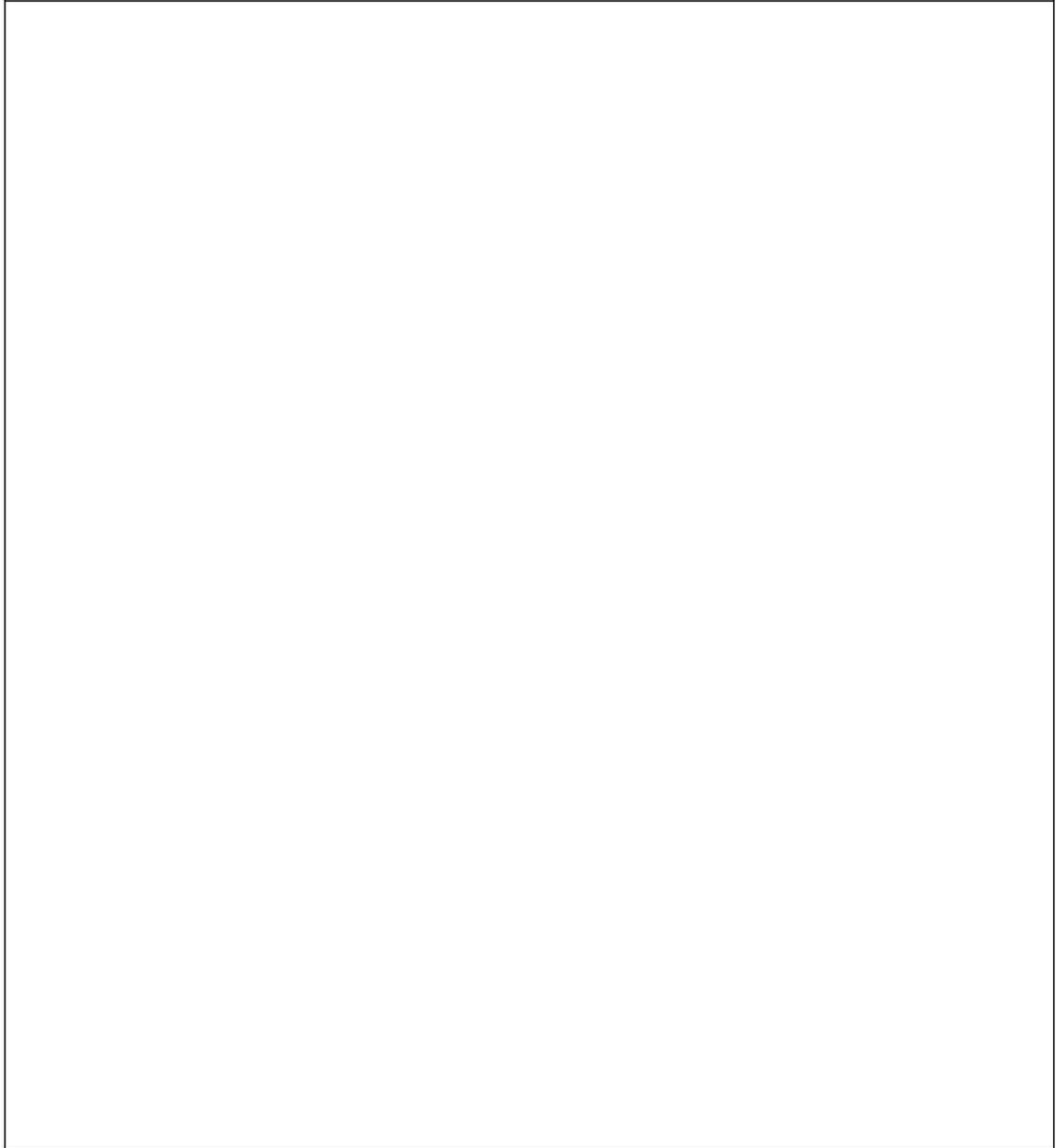
SURVEY ON DEVELOPING AN ASSESSMENT INSTRUMENT FOR PUBLIC UNIVERSITY PRESIDENTS						
PLEASE COMPLETE AND RETURN THIS QUESTIONNAIRE TO YOUR DEPARTMENT'S ADMINISTRATION OFFICE BY TBD USING THE ENVELOPE PROVIDED.						
Questions? Call Dennis Lester (505)853-1397 or e-mail to dlester@unm.edu						
The primary purpose of this survey is to obtain your opinion of the importance of various criteria that raters can use to assess the performance of a public university president. Additional purposes are to obtain feedback on the preferred approach to assessing performance and to identify factors that can bias the performance ratings given to a president. Its purpose is not to evaluate the performance of the current president of New Mexico State University or the University of New Mexico.						
<p>Instructions:</p> <ul style="list-style-type: none"> ▪ Using a pencil or pen, please place an X in the box that corresponds to how you would rate each item that follows. Example of a correct mark: ☒ ▪ With the exception of demographic items at the end of the survey, please mark only one box in response to each item. For the demographic questions, mark all applicable boxes. ▪ Consider the full range of the scale from "not important" to "critically important" in marking your choices so that significant differences may be identified during data analysis. ▪ If you are undecided, have no opinion, or do not want to respond to an item, please skip it and move to the next item. You may discontinue this survey at anytime. <p>Definitions: Constituents are members of the Board of Regents and faculty leadership. Stakeholders are all others who have an interest in the university.</p>						
Please rate the following items in terms of their importance as criteria for assessing the performance of a public university president.	Critically Important	5				
	Very Important	4				
	Important	3				
	Slightly Important	2				
	Not Important	1	1	2	3	4
1. Increases the quality of the faculty at the university						
2. Promotes institutional interests rather than self-interests						
3. Reduces the student-to-faculty ratio						
4. Maintains personal awareness of the climate within the institution						
5. Benchmarks with other institutions to identify improvement opportunities						
6. Makes responsible decisions regarding the allocation of resources						
7. Makes judicious decisions regarding the selection of senior administrators						
8. Promotes harmony among constituents/stakeholders						
9. Articulates the university story (e.g., vision, mission, values, and core competencies)						
10. Increases the quality of incoming students						
11. Clears obstacles that stand in the way of individual/institutional progress						
12. Adapts to changes that impact the university						
13. Shows appreciation to others for their support of the university						
14. Makes informed decisions based on best available information/research						
15. Provides clear direction/expectations when assigning tasks to the staff						
16. Rewards the superior performance of individuals/teams within the organization						
17. Serves as an inspiration to constituents/stakeholders						
18. Develops realistic plans to implement the university's strategy						
19. Maintains effective control of the budget						
20. Establishes effective teams to carry out specific tasks						
21. Recruits/retains employees that reflect the diversity of the State						
22. Builds trusting relationships with constituents/stakeholders						
23. Displays a high degree of job competence						
24. Promotes positive change within the institution						
25. Increases the student enrollment						
Proceed to item 26 on the next page						

Please rate the following items in terms of their importance as criteria for assessing the performance of a public university president.	Critically Important	5					
	Very Important	4					
	Important	3					
	Slightly Important	2					
	Not Important	1	1	2	3	4	5
26. Displays courage when faced with the challenges of university governance							
27. Promotes effective negotiation among interested parties to resolve conflicts							
28. Establishes realistic goals/objectives for the university							
29. Provides support to other leaders in the institution							
30. Performs benefit-risk analysis as part of the decision-making process							
31. Provides thoughtful responses to concerns expressed by constituents/stakeholders							
32. Uses appropriate performance indicators to monitor progress toward goals/objectives							
33. Promotes initiatives that contribute to the quality of education in New Mexico							
34. Incorporates lessons learned into university planning/operations							
35. Encourages open sharing of information with others							
36. Displays professional courtesy to others							
37. Provides a framework that aligns strategy/people to the mission of the university							
38. Increases the success of students							
39. Demonstrates a high degree of personal integrity							
40. Secures adequate resources to support the university's mission							
41. Recruits high caliber personnel to fill positions in the university							
42. Advocates for the university to strengthen its position/reputation in higher education							
Please rate the following items in terms of their importance as characteristics of an effective assessment process for a public university president.							
43. Provides feedback to the president for personal development/improvement							
44. Involves multiple constituents/stakeholders in the process (e.g., 360° feedback)							
45. Uses an assessment instrument that includes specific performance criteria							
46. Involves a systematic process that is executed on a regular basis (e.g., annually)							
47. Consists of a formal process with written policies, procedures, and assessment form							
48. Consists of an informal process without written guidelines or assessment form							
49. Ties president compensation to performance assessment ratings							
Please rate the following items in terms of their importance as factors that can bias the performance assessment of a public university president.							
50. Political pressure from state/local government officials							
51. Declining economy at national/state/local levels							
52. Conflicting priorities among constituents who participate in shared governance							
53. Scholarly criticism/skepticism of the constituents/stakeholders							
54. Rater errors (e.g., halo effect, leniency, and central tendency)							
55. Rater's limited knowledge of the roles/responsibilities of the president							
56. Rater's limited knowledge of the actions/decisions of the president							

Proceed to item 57 on the next page

Please provide the following demographic information. Mark an X in all applicable boxes.	
57. What is your age?	<input type="checkbox"/> 20 to 29 years <input type="checkbox"/> 30 to 39 years <input type="checkbox"/> 40 to 49 years <input type="checkbox"/> 50 to 59 years <input type="checkbox"/> 65 years or older
58. What is your gender?	<input type="checkbox"/> Male <input type="checkbox"/> Female
59. What is your race/ethnicity?	<input type="checkbox"/> American Indian <input type="checkbox"/> African American/Black <input type="checkbox"/> Chinese <input type="checkbox"/> Hispanic <input type="checkbox"/> Japanese <input type="checkbox"/> Korean <input type="checkbox"/> Vietnamese <input type="checkbox"/> White <input type="checkbox"/> Other, please specify: _____
60. What is your status as a faculty member at UNM?	<input type="checkbox"/> Full-time faculty <input type="checkbox"/> Part-time faculty <input type="checkbox"/> Assistant professor (tenure/tenure-track) <input type="checkbox"/> Associate professor (tenure/tenure-track) <input type="checkbox"/> Professor (tenure/tenure-track) <input type="checkbox"/> Instructional faculty <input type="checkbox"/> Visiting faculty <input type="checkbox"/> Clinician educator <input type="checkbox"/> Research faculty <input type="checkbox"/> Temporary faculty <input type="checkbox"/> Post-doctoral fellow <input type="checkbox"/> Other, please specify: _____
61. How many years experience do you have teaching at the college or university level?	<input type="checkbox"/> Less than 5 years <input type="checkbox"/> 5 to 9 years <input type="checkbox"/> 10 to 19 years <input type="checkbox"/> 20 or more years
62. How many years have you taught at your current university?	<input type="checkbox"/> Less than 5 years <input type="checkbox"/> 5 to 9 years <input type="checkbox"/> 10 to 19 years <input type="checkbox"/> 20 or more years
63. Where do you serve as a faculty member?	<input type="checkbox"/> New Mexico State University (4-year institution) <input type="checkbox"/> New Mexico State University (2-year institution) <input type="checkbox"/> University of New Mexico
64. Do you have additional comments on criteria for assessing university president success, best practices for assessing university presidents, factors that should be taken into account when assessing university presidents, and/or this survey? If so, please write them on the next page.	

Comments for Item 64



Thank you for completing this survey!

Please return your completed survey to your Department's Administrative Office no later than **TBD**.

Table E1. Final Survey Coding

Survey Item Number	Coded Descriptor	Original Theoretical/Conceptual Category
Traits, Behaviors, Performance Outcomes		
1	FacultyQuality	Performance Outcome Factor
2	InstitutionalInterests	Followership Factor
3	StudentFacultyRatio	Performance Outcome Factor
4	ClimateAwareness	Organization Factor
5	Benchmarking	Organization Factor
6	ResourceDecisions	Management Factor
7	AdministratorSelection	Management Factor
8	Harmony	Learning Factor
9	UniversityStory	Learning Factor
10	StudentQuality	Performance Outcome Factor
11	ClearObstacles	Leadership Factor
12	AdaptsToChange	Learning Factor
13	ShowsAppreciation	Followership Factor
14	InformedDecisions	Learning Factor
15	ClearDirection	Management Factor
16	RewardsPerformance	Learning Factor
17	Inspirational	Leadership Factor
18	RealisticPlans	Management Factor
19	ControlsBudget	Organization Factor
20	Teambuilding	Organization Factor
21	EmployeeDiversity	Performance Outcome Factor
22	BuildsTrust	Leadership Factor
23	JobCompetence	Leadership Factor
24	PromotesChange	Followership Factor
25	StudentEnrollment	Performance Outcome Factor
26	DisplaysCourage	Followership Factor
27	ResolvesConflicts	Followership Factor
28	SetsGoalsObjectives	Organization Factor
29	SupportsOtherLeaders	Followership Factor
30	PerformsRiskAnalysis	Management Factor
31	ThoughtfulResponses	Followership Factor
32	PerformanceIndicators	Management Factor
33	PromotesEducation	Organization Factor
34	LessonsLearned	Learning Factor
35	SharesInformation	Learning Factor
36	DisplaysCourtesy	Leadership Factor
37	StrategicFramework	Leadership Factor
38	StudentSuccess	Performance Outcome Factor
39	DisplaysIntegrity	Leadership Factor
40	SecuresResources	Performance Outcome Factor

Table E1 Continued

Survey Item Number	Coded Descriptor	Original Theoretical/Conceptual Category
41	RecruitsGoodPeople	Management Factor
42	UniversityAdvocate	Organization
Performance Assessment		
43	AssessFeedback	Performance Assessment Factor
44	Assess360degree	Performance Assessment Factor
45	AssessSpecificCriteria	Performance Assessment Factor
46	AssessRegularly	Performance Assessment Factor
47	AssessFormal	Performance Assessment Factor
48	AssessInformal	Performance Assessment Factor
49	AssessCompensation	Performance Assessment Factor
External Factors		
50	ExFacPolitics	External Factor
51	ExFacEconomy	External Factor
52	ExFacConflictingPriorities	External Factor
53	ExFacScholarlyCriticism	External Factor
54	ExFacRater Error	External Factor
55	ExFacKnowledgeofRoles	External Factor
56	ExFacKnowledgeofActions	External Factor

APPENDIX F
CONSENT FORMS

CONSENT FORM – INTERVIEW

Letter of Invitation and Consent

Developing an Effective Instrument for Assessing the Performance of Public University Presidents

- **INTRODUCTION**

You are invited to participate in a research study conducted by Dennis L. Lester, Ph.D. Candidate, from the Educational Leadership and Organizational Learning Department at the University of New Mexico (UNM). The results from this research project will contribute to a dissertation. You were identified as a possible volunteer in the study because you are a faculty member at UNM or New Mexico State University (NMSU).

- **PURPOSE OF THE STUDY**

The primary purpose of this study is to identify a process for developing the content and format of an effective performance assessment instrument for presidents of public universities and to identify candidate items to include in the assessment instrument. The results from this study will identify the characteristics of an assessment instrument that can be applied to conduct fair and meaningful performance assessments for public university presidents. The analysis and recommendations from this study will take into account the expectations of a key stakeholder group (i.e., the faculty); the strategy, goals, and objectives of universities; and the characteristics of performance assessment systems used at representative public universities. The purpose of this study is not to assess the performance of the current presidents of UNM or NMSU or the president of any other university.

- **PROCEDURES AND ACTIVITIES**

In order to participate in this study, you must be a faculty member at UNM or NMSU on a full-time or part-time basis. If you choose to participate, please attend the individual interview or focus group interview as specified in the enclosed letter of invitation. The purpose of this interview is to obtain faculty opinions is designed to obtain your opinion on traits, behaviors, performance outcomes, types of assessment, and external factors that may influence the performance of university presidents. You will also be asked to provide feedback on the emerging results of this study and the quality of the survey instrument.

- **POTENTIAL RISKS AND DISCOMFORTS**

This topic may be uncomfortable for faculty, particularly those who have a professional or personal relationship with senior administrators at UNM or NMSU such as members of the Board of Regents, staff in the Office of the President, or other individuals who have frequent contact with the UNM or NMSU president. If you choose to participate in this survey, you may stop at anytime and you do not have to answer all questions in the survey. If you have any questions about this study, you may contact Mr. Lester or the Institutional Review Board (IRB) for Research at UNM. Their contact information is contained in the Identification of Investigators and Review Board paragraph that follows.

- **POTENTIAL BENEFITS TO PARTICIPANTS AND/OR TO SOCIETY**

While the benefits of this study cannot be guaranteed, there is good potential that the results can be used to improve assessment of university presidents and to promote individual and organizational development. The Association of Governing Boards of Universities and Colleges (AGB) has conducted studies and produced reports on the characteristics of presidents of higher education institutions in the United States. The AGB has expressed interest in the results of this study. The Office of the President and Office of the Provost are being kept informed on the progress of this study.

- **CONFIDENTIALITY**

Any information obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Your responses in this interview will be consolidated with those responses from other participants and your name will not be associated with the results of this interview.

- **PARTICIPATION AND WITHDRAWAL**

You can choose to participate in this study or to not. If you volunteer to participate, you may withdraw at any time without penalty or loss of benefits to which you might otherwise be entitled. You may also refuse to answer any questions you do not want to answer and still remain in the study.

- **IDENTIFICATION OF INVESTIGATORS AND REVIEW BOARD**

If you have any questions or concerns about the research, please feel free to contact: Dennis Lester, MBA, 505-853-1397, dlester@unm.edu, or College of Education, ELOL, MSC05 3040, 1 University of New Mexico, Albuquerque, NM 87131-0001. If you have other concerns or complaints, contact the Institutional Review Board at the University of New Mexico, 1717 Roma NE, Room 205, Albuquerque, NM 87131, (505) 277-2257, or toll free at 1-866-844-9018.

- **STUDY RESULTS**

The results of this study will be summarized and included in an oral presentation and a dissertation manuscript. If you would like to see the final results of this study, please contact Dennis Lester. The final results will contain summary information that cannot be linked to individual responses.

Thank you for taking the time to participate in this study.

SIGNATURE OF RESEARCH PARTICIPANT

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been provided a copy of this form.

Name of Participant (please print)

Signature of Participant

Date

SIGNATURE OF INVESTIGATOR

In my judgment the participant is voluntarily and knowingly providing informed consent and possesses the legal capacity to give informed consent to participate in this research study

Name of Investigator or Designee

Signature of Investigator or Designee

Date

CONSENT FORM – SURVEY

Letter of Invitation and Consent

Developing an Effective Instrument for

Assessing the Performance of Public University Presidents

- **INTRODUCTION**

You are invited to participate in a research study conducted by a Ph.D. candidate, from the College of Education at the University of New Mexico (UNM). The results from this research project will contribute to a dissertation. You were identified as a possible volunteer in the study because you are a faculty member at UNM or New Mexico State University (NMSU).

- **PURPOSE OF THE STUDY**

The primary purpose of this study is to identify a process for developing the content and format of an effective performance assessment instrument for presidents of public universities and to identify candidate items to include in the assessment instrument. The results from this study will identify the characteristics of an assessment instrument that can be applied to conduct fair and meaningful performance assessments for public university presidents. The analysis and recommendations from this study will take into account the expectations of a key stakeholder group (i.e., the faculty); the strategy, goals, and objectives of universities; and the characteristics of performance assessment systems used at representative public universities. The purpose of this study is not to assess the performance of the current presidents of UNM or NMSU or the president of any other university.

- **PROCEDURES AND ACTIVITIES**

In order to participate in this study, you must be a faculty member at UNM or NMSU on a full-time or part-time basis. StudentVoice™ will administer a survey instrument via its website. The survey is designed to obtain your opinion on traits, behaviors, performance outcomes, types of assessment, and external factors that may influence the performance of university presidents. Your responses to this survey will be consolidated with those responses from other participants and your name will not be associated with the results of the survey. The survey will take approximately 15 minutes to complete.

- **POTENTIAL RISKS AND DISCOMFORTS**

This topic may be uncomfortable for faculty, particularly those who have a professional or personal relationship with senior administrators at UNM or NMSU such as members of the Board of Regents, staff in the Office of the President, or other individuals who have frequent contact with the UNM or NMSU president. If you choose to participate in this survey, you may stop at anytime and you do not have to answer all questions in the survey. If you have any questions about this study, you may contact the investigator or the Institutional Review Board (IRB) for Research at UNM. Their contact information is contained in the Identification of Investigators and Review Board paragraph that follows.

- **POTENTIAL BENEFITS TO PARTICIPANTS AND/OR TO SOCIETY**

While the benefits of this study cannot be guaranteed, there is potential that the results can be used to improve assessment of university presidents and to promote individual and organizational development. The Association of Governing Boards of Universities and Colleges (AGB) has conducted studies and produced reports on the characteristics of presidents of higher education institutions in the United States. The AGB has expressed interest in the results of this study. The Office of the President and Office of the Provost are being kept informed on the progress of this study.

- **CONFIDENTIALITY**

Any information obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. The survey administered via the StudentVoice website is completely anonymous.

- **PARTICIPATION AND WITHDRAWAL**

You can choose to participate in the survey or to not. If you volunteer to participate, you may withdraw at any time without penalty or loss of benefits to which you might otherwise be entitled. You may also refuse to answer any questions you do not want to answer and still remain in the survey.

- **IDENTIFICATION OF INVESTIGATORS AND REVIEW BOARD**

If you have any questions or concerns about the research, please feel free to contact: Dennis Lester, MBA, 505-853-1397, dlester@unm.edu, or College of Education, ELOL, MSC05 3040, 1 University of New Mexico, Albuquerque, NM 87131-0001. If you have other concerns or complaints, contact the Institutional Review Board at the University of New Mexico, 1717 Roma NE, Room 205, Albuquerque, NM 87131, (505) 277-2257, or toll free at 1-866-844-9018.

- **STUDY RESULTS**

The results of this study will be summarized and included in an oral presentation and a dissertation manuscript. If you would like to see the final results of this study, please contact Dennis Lester. The final results will contain summary information that cannot be linked to individual responses.

Thank you for taking the time to participate in this survey.

SIGNATURE OF RESEARCH PARTICIPANT

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been provided a copy of this form.

Name of Participant (please print)

Signature of Participant

Date

SIGNATURE OF INVESTIGATOR

In my judgment the participant is voluntarily and knowingly providing informed consent and possesses the legal capacity to give informed consent to participate in this research study

Name of Investigator or Designee

Signature of Investigator or Designee

Date

APPENDIX G
DATA COLLECTION FORMS

GROUNDING THEORY CODING WORKSHEET

Researcher Name:

Page ____ of

Coding (circle applicable code)

Interviewee Comments (circle applicable code)

Initial: categories/in vivo
Axial: categories/subcategories/
properties/dimensions
Theoretical:
categories/subcategories

Initial: all comments
Axial: focused comments
Theoretical: relationships and hypotheses

--	--

APPENDIX H
INSTITUTIONAL REVIEW BOARD (IRB) APPROVALS



Office of the Vice President for Research

INSTITUTIONAL REVIEW BOARD (IRB)
Dr. Luis A. Vazquez, Chair

MSC 3RES-PSL
New Mexico State University
P. O. Box 30001
Las Cruces, NM 88003-8001
Phone: 575-646-7177 Fax: 575-646-2480
Email: ovpr@nmsu.edu

DATE: September 16, 2009
TO: Dennis L. Lester
FROM: Nellie Quezada-Aragon *Nellie Quezada-Aragon*
SUBJECT: Application for Permission to Use Human Subjects in Research
IRB Application Number: **325 (Exempt) Revision 01**

The NMSU Institutional Review Board Chair, Dr. Luis A. Vazquez, has reviewed your request for an extension of approval with the changes indicated in the request for modification submitted for the project titled, **“Developing an Effective Instrument for Assessing the Performance of Public University Presidents.”**

The application was reviewed in accordance with the expedited review process outlined in 45 CFR 46.110(b)(1) - Category 7. Dr. Vazquez has determined your project to be exempt from full IRB review according to 45 CFR Part 46.101(b)(2) and (b)(4).

Dr. Vazquez approved the application on behalf of the IRB on September 11, 2009. Your IRB approval is valid for the period: **September 11, 2009 - August 25, 2010.**

The research must be conducted according to the proposal/protocol that was approved by the IRB. Any changes in the research proposal, instruments, or the consent document(s) must be submitted to the IRB prior to implementation. Additionally, any unexpected hazards or adverse events involving risk to the subjects or others must be reported immediately to the IRB.

Please note that the IRB approval is valid for only one (1) year. The IRB must review and approve all research protocols involving human subjects at intervals appropriate to the degree of risk, but not less than once per year. Therefore, in order to continue your project after the approved period, you must submit a request for continuation **60 days prior** to the end date of **August 25, 2010.**

If you should have any questions, please do not hesitate to contact me at 646-7177 or via e-mail at [<ovpr@nmsu.edu>](mailto:ovpr@nmsu.edu).

cc: Dr. Luis A. Vazquez, IRB Chair



THE UNIVERSITY *of*
NEW MEXICO

Main Campus Institutional Review Board

Human Research Protections Office

1717 Roma NE, MSC05 3180

1 University of New Mexico-Albuquerque, NM 87131-0001

<http://hsc.unm.edu/son/research/HRRC/>

22-Jan-2009

Responsible Faculty: Patsy Boverie

Investigator: Dennis L. Lester

Dept/College: Educ Leadership Orgn Learning ELOL

SUBJECT: IRB Determination of Exempt Status

Protocol #: 08-607

Project Title: Developing an Effective Instrument for Assessing the Performance of Public University Presidents

Approval Date: 12-Jan-2009

The Main Campus Institutional Review Board has reviewed the above-mentioned research protocol and determined that the research is EXEMPT from the requirements of Department of Health and Human Services (DHHS) regulations for the protection of human subjects as defined in 45CFR46.101(b) under category 1, and the Food and Drug Administration (FDA) regulations as defined in 21CFR50.1 and 21CFR56.101 do not apply to research. Therefore, this research project is not subject to continuing review.

Changes to the Research: It is the responsibility of the Principal Investigator to inform the IRB of any changes to this research. A change in the research may disqualify this project from exempt status. Reference the protocol number and title in all documents related to this protocol.

Sincerely,

J. Scott Tonigan, PhD

Chair

Main Campus IRB



THE UNIVERSITY *of*
NEW MEXICO

Main Campus Institutional Review Board

Human Research Protections Office

1717 Roma NE, MSC05 3180

1 University of New Mexico-Albuquerque, NM 87131-0001

<http://hsc.unm.edu/som/research/HRRC/>

11-Mar-2009

Responsible Faculty: Patsy Boverie

Investigator: Dennis L. Lester

Dept/College: Educ Leadership Orgn Learning ELOL

SUBJECT: IRB Approval of Research - Amendment

Protocol #: 08-607

Project Title: Developing an Effective Instrument for Assessing the Performance of Public University Presidents

Type of Review: Exempt Review

Approval Date:

Expiration Date:

The Main Campus Institutional Review Board has reviewed and approved the above referenced protocol. It has been approved based on the review of the following:

1. UNM Consent version 02/28/09;
2. Letter of invitation submitted 03/02/09.

Consent Decision:

Amended consent(s) attached.

When consent is required, it is the responsibility of the Principal Investigator (PI) to ensure that ethical and legal informed consent has been obtained from all research participants. A date stamped original of the approved consent form(s) is attached, and copies should be used for consenting participants during the above noted approval period.

As the principal investigator of this study, you assume the following responsibilities:

Renewal: Unless granted exemption, your protocol must be re-approved each year in order to continue the research. You must submit a Progress Report no later than 30 days prior to the expiration date noted above.

Adverse Events: Any adverse events or reactions must be reported to the IRB immediately.

Modifications: Any changes to the protocol, such as procedures, consent/assent forms, addition of subjects, or study design must be submitted to the IRB for review and approval.



THE UNIVERSITY *of*
NEW MEXICO

Main Campus Institutional Review Board
Human Research Protections Office
MSC08 4560

1 University of New Mexico-Albuquerque, NM 87131-0001
<http://hsc.unm.edu/som/research/HRRC/>

21-May-2009

Patsy Boverie PhD
Educ Leadership Orgn Learning ELOL
Dennis L. Lester

SUBJECT: IRB Approval of Amendment Change
HRRC#: 08-607
Study Title: Developing an Effective Instrument for Assessing the Performance of Public University Presidents
Type of Review: Exempt Review
Approval Date: 21-May-2009

Dear Dennis L. Lester:

The Institutional Review Board (IRB) has reviewed and approved the following amendment(s) and/or change(s) to the above mentioned protocol:

1. Update to protocol dated 04/13/09;
2. Interview Consent version 04/13/09;
3. Survey Consent version 04/13/09;
4. Focus Group Interview Cover Letter submitted 04/13/09;
5. Individual Interview Cover Letter submitted 04/13/09;
6. Dissertation presentation submitted 04/13/09;
7. Survey Invitation Cover Letter submitted 04/13/09;
8. Interview Guides submitted 04/13/09.

This study continues to be exempt from annual review with the UNM IRB.

This research is still EXEMPT from the requirements of Department of Health and Human Services (DHHS) regulations for the protection of human subjects as defined in 45CFR46.101(b). Therefore, this research project is not subject to IRB continuing review.

Changes to the Research: It is the responsibility of the Principal Investigator and/or Responsible Faculty to inform the IRB of any changes to this research. A change in the research may disqualify this project from exempt status. Reference the IRB# and title in all documents related to this protocol.

Sincerely,



THE UNIVERSITY *of*
NEW MEXICO

Main Campus Institutional Review Board
Human Research Protections Office
MSC08 4560

1 University of New Mexico-Albuquerque, NM 87131-0001
<http://hsc.unm.edu/som/research/HRRC/>

12-Aug-2009

Responsible Faculty: Patsy Boverie
Investigator: Dennis L. Lester
Dept/College: Educ Leadership Orgn Learning ELOL

SUBJECT: IRB Approval of Research - Amendment
Protocol #: 08-607

Project Title: Developing an Effective Instrument for Assessing the Performance of Public University Presidents

Type of Review: Exempt Review

Approval Date: 11-Aug-2009

Expiration Date:

The Main Campus Institutional Review Board has reviewed and approved the above referenced protocol. It has been approved based on the review of the following:

1. Protocol submitted 07/02/09 (exempt);
2. Interview Consent version 07/02/09;
3. Survey Consent version 07/02/09;
4. Survey on Developing an Assessment Instrument for Public University Presidents submitted 07/02/09;
5. Letter of Invitation for focus group interview submitted 07/09/09;
6. Letter of Invitation for final survey submitted 07/09/09;
7. Letter of Invitation for personal interview submitted 07/09/09.

Consent Decision:
Amended consent(s) attached.

When consent is required, it is the responsibility of the Principal Investigator (PI) to ensure that ethical and legal informed consent has been obtained from all research participants. A date stamped original of the approved consent form(s) is attached, and copies should be used for consenting participants during the above noted approval period.

As the principal investigator of this study, you assume the following responsibilities:

Renewal: Unless granted exemption, your protocol must be re-approved each year in order to continue the research. You must submit a Progress Report no later than 30 days prior to the expiration date noted above.



THE UNIVERSITY *of*
NEW MEXICO

Main Campus Institutional Review Board

Human Research Protections Office

MSC08 4560

1 University of New Mexico-Albuquerque, NM 87131-0001

<http://hsc.unm.edu/som/research/HRRC/>

08-Sep-2009

Responsible Faculty: Patsy Boverie

Investigator: Dennis L. Lester

Dept/College: Educ Leadership Orgn Learning ELOL

SUBJECT: IRB Approval of Research - Amendment

Protocol #: 08-607

Project Title: Developing an Effective Instrument for Assessing the Performance of Public University Presidents

Type of Review: Exempt Review

Approval Date: 04-Sep-2009

Expiration Date:

The Main Campus Institutional Review Board has reviewed and approved the above referenced protocol. It has been approved based on the review of the following:

Survey on Developing an Assessment Instrument for Public University Presidents submitted 09/01/09.

Consent Decision:

No changes.

When consent is required, it is the responsibility of the Principal Investigator (PI) to ensure that ethical and legal informed consent has been obtained from all research participants. A date stamped original of the approved consent form(s) is attached, and copies should be used for consenting participants during the above noted approval period.

As the principal investigator of this study, you assume the following responsibilities:

Renewal: Unless granted exemption, your protocol must be re-approved each year in order to continue the research. You must submit a Progress Report no later than 30 days prior to the expiration date noted above.

Adverse Events: Any adverse events or reactions must be reported to the IRB immediately.

Modifications: Any changes to the protocol, such as procedures, consent/assent forms, addition of subjects, or study design must be submitted to the IRB for review and approval.

Completion: When the study is concluded and all data has been de-identified (with no link to identifiers),



THE UNIVERSITY *of*
NEW MEXICO

Main Campus Institutional Review Board

Human Research Protections Office

MSC08 4560

1 University of New Mexico-Albuquerque, NM 87131-0001

<http://hsc.unm.edu/som/research/HRRC/>

15-Sep-2009

Patsy Boverie PhD
Educ Leadership Orgn Learning ELOL
Dennis L. Lester

SUBJECT: IRB Approval of Amendment Change
HRRC#: 08-607
Study Title: Developing an Effective Instrument for Assessing the Performance of Public University Presidents
Type of Review: Exempt Review
Approval Date: 14-Sep-2009

Dear Dennis L. Lester:

The Institutional Review Board (IRB) has reviewed and approved the following amendment(s) and/or change(s) to the above mentioned protocol:

Investigator Protocol version 08/31/09, including increase in target to 500 subjects.

This research is still EXEMPT from the requirements of Department of Health and Human Services (DHHS) regulations for the protection of human subjects as defined in 45CFR46.101(b). Therefore, this research project is not subject to IRB continuing review.

Changes to the Research: It is the responsibility of the Principal Investigator and/or Responsible Faculty to inform the IRB of any changes to this research. A change in the research may disqualify this project from exempt status. Reference the IRB# and title in all documents related to this protocol.

Sincerely,

J. Scott Tonigan, PhD
Chair
Main Campus IRB

* Under the provisions of this institution's Federal Wide Assurance (FWA00004690), the Main Campus IRB has determined that this proposal provides adequate safeguards for protecting the rights and welfare of the subjects involved in the study and is in compliance with HHS Regulations (45 CFR 46).



THE UNIVERSITY *of*
NEW MEXICO

Main Campus Institutional Review Board

Human Research Protections Office

MSC08 4560

1 University of New Mexico-Albuquerque, NM 87131-0001

<http://hsc.unm.edu/som/research/HRRC/>

03-Dec-2009

Patsy Boverie PhD
Educ Leadership Orgn Learning ELOL
Dennis L. Lester

SUBJECT: IRB Approval of Amendment Change
HRRC#: 08-607
Study Title: Developing an Effective Instrument for Assessing the Performance of Public University Presidents
Type of Review: Exempt Review
Approval Date: 02-Dec-2009

Dear Dennis L. Lester:

The Institutional Review Board (IRB) has reviewed and approved the following amendment(s) and/or change(s) to the above mentioned protocol:

Investigator Protocol submitted November 11, 2009.

This research is still EXEMPT from the requirements of Department of Health and Human Services (DHHS) regulations for the protection of human subjects as defined in 45CFR46.101(b). Therefore, this research project is not subject to IRB continuing review.

Changes to the Research: It is the responsibility of the Principal Investigator and/or Responsible Faculty to inform the IRB of any changes to this research. A change in the research may disqualify this project from exempt status. Reference the IRB# and title in all documents related to this protocol.

Sincerely,

J. Scott Tonigan, PhD
Chair
Main Campus IRB

* Under the provisions of this institution's Federal Wide Assurance (FWA00004690), the Main Campus IRB has determined that this proposal provides adequate safeguards for protecting the rights and welfare of the subjects involved in the study and is in compliance with HHS Regulations (45 CFR 46).

APPENDIX I
INITIAL INTERVIEW FINDINGS

Tables I-1 through I-8 in this appendix provide a summary of the feedback from the initial individual interviews and the focus group interviews. The number of marks in the columns “individual interviews” and “focus group interviews” indicates how many interviewees provided the same responses to an interview question. A single mark in a column indicates that only one interviewee provided the response. See appendix B and C for complete statements of the questions posed to the interviewees. The order of the tables with the corresponding questions is as follows:

- Table I-1: What are the traits of a good university president?
- Table I-2: What are the positive behaviors of a university president?
- Table I-3: What are the negative behaviors of a university president?
- Table I-4: What are the more relevant areas in which to assess presidential performance outcomes?
- Table I-5: What factors are beyond a president’s control that can affect his or her performance ratings
- Table I-6: Is formal assessment of value to the president and university?
(Note: Formal assessment includes written policies, procedures, assessment criteria, and administration of the assessment instrument on a regular basis.)
- Table I-7: What is the purpose or aim of university president assessment?
- Table I-8: Who should be involved in university president assessment?

Table I-1. Traits of a good university president

Traits	Individual Interview	Focus Group Interview
Honesty		
Leadership		
Interest in national education issues		
Commitment to excellence		
Academic vision		
Competence in facilitation		
Ability to translate vision into results		
Ability to partner with constituents		
Scholarship		
Integrity		
Institutional knowledge		
Effective communication		
Empathy		
Respect		
Persuasiveness		
Represents university well		
Focus		
Trust/trustworthiness		
Optimism		
Authenticity		
Skill at interpersonal relationships		
Courage		
Decision making ability		
Service-mindedness		
Academic skill		
Listening skills		
Ability to accept criticism		
Commitment to educating the community		
Ability to handle pressure		
Knowledge of research		
Strength of character		
Listening and learning skills		
Knowledge of human capital		
Knowledge of core mission		
Strong sense of purpose		
Dedication to public service		
Dedication to common agenda		
Strong faculty credentials		
Charisma		
“Down-to-earth”		
Graciousness		
Knowledge of university culture		
Ability to motivate		
Ability to inspire		
Knowledge of higher education		
Knowledge of university place in U.S. higher education		
Administrative skills		
Support of university constituents/stakeholders		

Traits	Individual Interview	Focus Group Interview
Efficiency		
Straight-forwardness		
Sound/prudential judgment		
Knowledge of internal/external influences		
Perceptiveness of influences on university		
“Healer/comforter”		
“Hands-off” leadership		
Credibility with constituents		
Perceptiveness of other people		
Management skills		
Humility		
Patience		
Vision (“sees the way ahead”)		
Openness/responsiveness to feedback		
Ability to expanded university influence		
Social intelligence		
Emotional intelligence		
Political skills		
Broad range of skills		
Ability to adapt to the environment		
Willingness to reach consensus		
Diplomacy		

Table I-2. Positive behaviors of a university president

Positive Behaviors	Individual Interview	Focus Group Interview
Shows appreciation/respect		
Knows individuals/audiences		
Has organization's best interests in mind		
Argues the university's case to the state		
Remains visible on and off campus		
Talks to people		
Takes responsibility for health/welfare of the organization		
Helps faculty do their jobs		
Encourages equal split in responsibility for goals/objectives		
Discerns if actions consistent with core vision/mission		
Allocates resources to core mission		
Identifies opportunities to advance the institution		
Explains decisions		
Serves as a mentor		
Demonstrates ability to recruit/appoint right people		
Demonstrates courage/stands up to pressure/adversity		
Shows willingness to sacrifice self-interests		
Demonstrates accessibility/transparency		
Makes cogent arguments		
Involves faculty at the strategic level		
Gets involved in the culture/state		
Demonstrates the ability to create a shared vision		
Builds coherent goals/objectives from the departments up		
Engages across campus		
Avoids easy way out		
Asks questions about what's going on in departments		
Relates lessons learned from other organizations		
Demonstrates approachability		
Cares about peoples' jobs		
Balances time with constituents		
Listens		
Demonstrates quickness on feet		
Shows command of issues		
Prepares for meetings		
Displays empathy		
Demonstrates ability to solve problems		
Taps internal talent		
Solicits feedback		
Demonstrates ability to determine what is relevant		
Demonstrates ability to mediate disputes		
Takes actions match words		
Talks-up accomplishments		
Acts responsibly		
Recognizes people		
Shows a sense of honor to serve		

Positive Behaviors	Individual Interview	Focus Group Interview
Endorses academic freedom		
Serves as a beacon for the university		
Demonstrates ability to persuade		
Takes actions that encourage loyalty		
Demonstrates ability to identify problems that can be addressed		
Demonstrates ability to deal with crises		
Serves as a lawyer, fixer, and convincer		
Moves with agility between constituents		
Demonstrates ability to read people		
Shows patience and persistence		
Discerns times to “nudge” and to “wait” on actions		
Involves constituents in vicarious learning		
Avoids personalizing issues, “it’s not about me”		
Demonstrates ability to win confidence		
Adapts role to serve institutional needs		
Demonstrates ability to have meaningful dialogue		
Unburdens faculty in day-to-day management		
Gives faculty strong voice increasing their participation		
Displays honesty		
Displays integrity		
Maintains self-awareness		
Maintains self-control		
Serves as an effective conflict manager		
Considers other peoples’ ideas		
Takes responsibility to prepare students to enter college		
Partners with other universities		
Has a plan and articulates it well		
Uses “SMART” objectives that tie to goals		

Table I-3. Negative behaviors of a university president

Negative Behaviors	Individual Interview	Focus Group Interview
Demonstrates a lack of transparency		
Acts dishonestly/deceitfully		
Makes patronage appointments		
Marginalizes/demeans shared governance		
Does not listen		
Allows politics to have undue influence		
Creates unnecessary administrative positions		
Diverts academic funding to administration		
Demonstrates self-centeredness/ties to own vision		
Does not take inputs before making decisions		
Acts defensively/takes things personally		
Acts aloof/shows lack connectedness		
Tells people what to do		
Lacks respect		
Does not visit academic departments or get around campus		
Adopts corporate business model		
Ignores advice		
Has too many goals upon which to act		
Justifies decisions by “this is what I said I would do”		
Does not gather adequate information to defend position		
Lacks perspective		
Exhibits disdain for individuals		
Does not meet with faculty		
Authorizes big bonuses/salaries/severance packages		
Makes excuses		
Makes pronouncements without personal contact		
Lacks knowledge of audiences		
Demonstrates an inability to resolve issues		
Displays impulsiveness		
Does unethical things		
Lacks allegiance/loyalty		
Does things that dishonor the university		
Recruits expensive faculty		
“Knuckles under to people”		
Over-centralizes/acts autocratically		
Seeks public attention for personal ambitions		
Demeans people		
Hires poor quality people		
Does not follow approved procedures		
Takes out frustrations on others		
Provides canned responses to issues		
Blows things out of proportion		
Demonstrates arrogance (stemming from insecurity)		
Demonstrates unresponsiveness		
Rushes the decision-making process		
Does not taking faculty seriously		
Manipulates data to tell desired story		

Negative Behaviors	Individual Interview	Focus Group Interview
Lacks clarity on issues		
Takes retribution against nonconformists		
Makes assumptions about peoples' opinions/perspectives		
Creates filters so people say only what you want to hear		
Demonstrates inability to generate resources		

Table I-4. Preferred performance outcome measures

Performance Outcomes	Individual Interview	Focus Group Interview
Mood/climate/morale of faculty/organization		
Resource acquisition		
University structure/systems development		
Support of core mission		
Contributions to endowment		
Initiatives that make a difference		
Responsiveness to crises		
Progress toward building a higher quality faculty		
Quality of life		
Quality of education		
Student-to-faculty ratio		
Student success		
Indicators that show movement in the right direction		
Hiring practices		
Resource allocation (to core mission of teaching/research)		
UNM Dashboard indicators (with faculty/constituent inputs)		
General and specific goals of the university		
Progress toward goals/objectives		
Reputation of departments		
Diversity of faculty		
Stability of faculty		
Engagement with faculty		
Engagement with constituents		
Faculty success (grants, teaching awards, Fulbright scholarships)		
Problem solving		
Financial status		
Enrollment		
Quality of students		
Graduation rates		
Percent of undergraduates who enter graduate school		
Scholarships		
University reputation (in multiple dimensions)		
Progress in areas such as diversity		
Actions on special issues (e.g., non-resident campus)		
Generation of new ideas		
Balance between symbolic and substantive decisions		
Items generated by using a balanced scorecard approach		
Creation of meaning and purpose in actions		
Performance predictors that may correlate with longer-term outcomes such as student retention		
Trends in various performance outcomes		
Progress toward initiatives that capitalize on university strengths		
Outcomes that have targeted thresholds		

Performance Outcomes	Individual Interview	Focus Group Interview
Measures that take into account the era, economy, purpose, and mission of the university	1	
Performance measures in which the outcomes cannot be manipulated	1	

Table I-5. Factors beyond president's control that can affect performance ratings

External Factors	Individual Interview	Focus Group Interview
Economy*		
Regents' priorities*		
Quality of incoming students*		
Natural disasters*		
Decentralization/diffusion of authority with shared governance		
Politics/political pressure		
Funding from legislature		
Funding from endowments		
Social unrest		
Student enrollment		
Governor priorities		
Hidden agendas		
Critical audiences (e.g., faculty) whose job is to critique, analyze, and research		
Authority not being commensurate with accountability		
Resource limitations		
Amount of real power		
Reluctance followers		
Demographics of constituents/stakeholders		
Business community attitudes		
Environment external to the university		
Scholarly publication output		
Culture of the university		
Conflicting personalities		
Use of performance assessment results		
Success of athletic teams		
<p>*Note: While the president cannot control these factors, four interviewees said that the president is accountable for how he or she responds to these external factors. One interviewee said that trust relationships could compensate for factors beyond the president's control. Two interviewees said that the president could affect the quality of incoming students given an investment in time and effort so this is not an uncontrollable factor.</p>		

Table I-6. Value of formal assessment to the president and university

Value of Formal Assessment and Caveats	Individual Interview	Focus Group Interview
It is of value		
It is of value if made public		
It is of value if it involves the right people		
It is of value and good leaders will take the assessment to heart		
It is of value, but one must have confidence in the board of regents		
It is of value if president is interested in making changes		
It is absolutely of value		
It needs to be anonymous		
It should be a 360-degree assessment		
It should involve multiple constituents		
It should have quantitative and qualitative aspects		
It depends on the purpose		
It increases self-awareness		
Its usefulness depends how it is accomplished		
It could be instructive and helpful if university environment is conducive		
It should be conducted in conjunction with a climate survey		
Its formality should be left to the discretion of the president based on preferred means of feedback		
It provides visibility in blind spots		
It depends on the university environment		
It is crucial		
It is valuable before renewal of any contract		
It may not be of any value, it could be used as a club		
It may provide too many perspectives to be useful		
It is important because it is the only way to provide accurate feedback		
It depends whether the results are used to make a difference		
It costs more and may be a bigger distraction		
Its strengths are people know what's happening and it has fixed procedures		
It would be valuable if formal assessment does not work		
It provides a calibration of self-concept		
It is not valuable if open communications already exist		
It should be a written with short declarative statements		
It should be conducted by experts to give credible feedback and to explain results to the regents		
It should be run by a committee		
It cannot be pro forma (i.e., square-filling)		
It must be open and transparent		
It must be used in the proper way		
It could be useful if there is an appropriate balance of power between the president and regents		

Value of Formal Assessment and Caveats	Individual Interview	Focus Group Interview
It should be responsive to the faculty		
It should be done as a group approach		
It depends on the leadership structure and whether the right people are involved		
It should construct by the regents with constituent inputs		
Its instrument should contain some previously defined categories and some narrative		
It should be done annually		
It is of no value of the instrument is not reliable/valid		
It would be counterproductive if the regents enforce/impose their views		
Its value depends on the role of shared governance		
Its value depends on culture and politics		
It is of no value if only the regents get the results		
It should be run by the faculty because they are most representative of the university		

Table I-7. Purpose of university president assessment

Purpose	Individual Interview	Focus Group Interview
Provide inputs for improvement		
Assess success in core mission (teaching, research, service, economic development, and patient services)		
Enable the president to operate more efficiently/successfully		
Provide feedback on how well doing		
Support retention decisions		
Set the tone for the university		
Provide accountability		
Evaluate performance of the president		
Support personal development		
Provide transparency/understanding of what the president is doing to help		
Nurture the university mission		
Evaluate resource allocation		
Identify things president needs to do to be successful		
Provide justification for merit pay increases		
Provide information to the regents		
Assess effectiveness of the president		
Provide emphasis external factors that affect performance		
Provide a tool for the president		
Increase understanding of what is expected by stakeholders		

Table I-8. Recommended participants in university president assessment

Participants	Individual Interview	Focus Group Interview
Board of regents		
Students		
Administrators		
Donors		
Constituents of president		
Faculty governance		
Alumni/alumni board		
Legislators		
Deans		
Faculty		
Student leadership		
Faculty senate		
Staff (fear of retribution unless anonymous)		
Vice presidents		
Peer presidents		
President of the board of regents		
Faculty senate president		
Emeritus faculty		
360-degree feedback (supervisors, peers, subordinates)		
Department chairs		
State secretary of higher education (possibly)		
Governor (possibly)		
President (self-assessment)		
Faculty council		
Regents' professors		
Community leaders		
Past presidents		
Faculty senate operations committee		
Legislators, donors, and alumni (uncertain of value of other participants)		
Academic freedom and tenure committee		
Associated Students of UNM (ASUNM)		
UNM Graduate and Professional Student Association (GPSA)		
Head of staff council		
Alumni board president		
Faculty committees		

APPENDIX J
PILOT SURVEY FINDINGS

This appendix provides a summary of the responses to the pilot survey. Table J1 contains the number of respondents (N), means, and standard deviations for each item in the pilot survey (see appendix D for exact statements of the items contained in the pilot survey). The items are arranged in the order from highest mean scores to lowest mean scores.

Table J-1. Pilot survey descriptive statistics

Items	N	Mean	Standard Deviation
Displays a high degree of personal integrity	106	4.75	.715
Promotes institutional interests rather than self-interests	100	4.71	.701
Builds trusting relationships with others	105	4.63	.724
Encourages open sharing of knowledge among constituents	106	4.58	.792
Displays a high degree of job competence	106	4.56	.744
Adapts to changes that affect the university	105	4.35	.796
Takes appropriate actions to secure resources	100	4.34	.768
Demonstrates professional courtesy to others	100	4.33	.888
Clears obstacles that enable constituents to be successful	104	4.33	.875
Student/faculty ratio	96	4.29	.845
Provides support to those in leadership roles	100	4.29	.756
Recruits high caliber personnel	100	4.29	.800
Articulates the university story (e.g., vision, mission, values)	106	4.24	.879
Incorporates lessons learned into decision-making	106	4.23	.908
Assessment process involves multiple stakeholders	99	4.22	.975
Establishes realistic goals for the institution	99	4.18	.837
Develops realistic plans to implement strategy	100	4.18	.869
Shared governance can influence president performance ratings	93	4.15	.859
Total revenue for the institution	96	4.15	.808
Conducts assessment on a regularly scheduled basis	100	4.04	1.024
Funding from state appropriations	96	4.03	.852
Displays passion toward his or her work	105	4.03	.904
Uses appropriate performance indicators to make decision	99	4.00	.969
Maintains good awareness of stakeholder satisfaction	100	3.99	.882
Provides a framework for developing institutional strategy	106	3.98	1.095
Serves as an agent for positive change	100	3.98	.829
Rewards superior performance	106	3.97	.980
Amount of research funding	96	3.97	.978
Percentage of full-time faculty	96	3.97	1.031
Level of satisfaction among stakeholders	96	3.96	.893
Student graduation rates	96	3.96	.905
Displays a willingness to reach consensus	100	3.95	1.009
Quality of campus life	96	3.94	.904
Maintains effective control over resources	99	3.93	.848
Establishes effective teams	100	3.92	.950

Items	<i>N</i>	Mean	Standard Deviation
Average number of students in a class	96	3.92	1.023
Applies previous knowledge/experience to solve problems	106	3.91	.911
Faculty compensation	95	3.89	.905
Student retention rate	96	3.89	.972
Percentage of classes taught by tenured faculty	95	3.87	1.074
Investment in facility modernization	96	3.82	.894
Promotes harmony among university stakeholders	106	3.81	1.079
Amount of foundation gifts	96	3.80	.902
Assessment system has documented policy/processes/procedures	99	3.80	.990
Assessment system includes an instrument with specific criteria	100	3.78	1.060
Ratings by peer institutions	95	3.77	1.026
Inspires constituents to follow his or her lead	106	3.76	1.092
Assessment system provides ongoing feedback for personal development	98	3.76	1.016
Follower perspectives of good leader can influence president performance ratings	93	3.73	.946
Multiple/conflicting goals of the university can influence president performance ratings	93	3.70	.857
Makes logical decisions regarding work assignments for staff	99	3.70	1.044
Amount of grant funding	96	3.69	.988
Involvement in the local community	96	3.69	.910
Provides clear direction for task execution	100	3.68	1.014
Assessment system includes objective vs. subjective criteria	99	3.68	1.067
Scholarly criticism/skepticism can influence president performance ratings	93	3.67	.925
Empowers individuals by decentralizing authority	99	3.64	1.191
Assessment system ties president compensation to president performance	99	3.64	1.165
Maintains an effective knowledge management system	99	3.61	.967
Amount of scholarship aid to students	96	3.60	1.010
Raises controversial issues	100	3.60	.816
Manages risk	99	3.59	.869
Promotes continuous learning for continuous improvement	100	3.58	1.148
Promotes initiatives that contribute to society	100	3.55	.999
Promotes self-reflection to transform old ways of thinking	106	3.55	1.139
Demonstrates assertiveness in resolving issues	105	3.54	.821
Placement of graduating students	96	3.53	1.085
Number of scholarly publications by faculty	96	3.51	1.046
Percentage of faculty with top terminal degrees	96	3.51	1.133
Incorporates a holistic approach to problem solving	106	3.50	1.197
Maintains an upbeat attitude	106	3.49	.959
Performs benchmarking to identify improvements	100	3.49	.980
Investment in advanced information technologies	96	3.47	1.025
Experience of stakeholders involved in an assessment can influence president performance ratings	92	3.47	.966
Assists individuals in achieving goals/aspirations	105	3.46	1.029
Tuition as a percentage of revenue	95	3.42	1.048

Items	<i>N</i>	Mean	Standard Deviation
Economic conditions surrounding the university can influence president performance ratings	93	3.37	.870
Percentage of minority faculty	96	3.33	1.053
Balance in the consolidated investment fund	94	3.32	1.090
Cost per semester hour	94	3.30	1.096
Suggests novel ways of doing business	106	3.29	.925
Number of full-time students	95	3.21	1.129
Emphasizes customer satisfaction	99	3.20	1.116
Amount of bequests received	94	3.19	1.029
Percentage of minority students	96	3.19	1.009
Student enrollment	96	3.17	1.149
Number of faculty awards	96	3.09	1.067
Cultural backgrounds of university stakeholders can influence president performance ratings	93	3.08	1.086
Number of degrees granted	96	3.04	.983
Focuses strategies on increased competitiveness	100	2.99	1.059
Student semester credit hours	95	2.95	1.035
Number of student awards	96	2.88	.897
Tailors leadership style to follower expectations	105	2.79	1.222
Rater errors can influence president performance ratings	91	2.75	.995
Number of Fulbright scholars	95	2.69	1.001
Availability of extracurricular activities	96	2.68	.923
Variables in which the president has no control can influence president performance ratings	93	2.66	.994
High school grade point average of incoming freshman	96	2.65	1.005
Class standing of incoming students	95	2.57	1.155
Number of transfers from other institutions	96	2.56	.960
Examinations scores of incoming students	96	2.55	1.045
Rate of participation in extracurricular activities	96	2.55	.928
Stakeholder desire to maintain autonomy from the staff can influence president performance ratings	92	2.47	1.074
Number of international students	96	2.35	.973
Number of patents issued	95	2.24	.884
Acceptance rate of incoming students	96	2.22	1.007
Number of license/option agreements	94	2.21	.890
Number of students studying abroad	96	2.17	.914
Number of start-up companies	94	2.09	.912
Success in intercollegiate athletics	96	1.80	1.042
Total revenue from athletic events	96	1.78	.997

APPENDIX K
FINAL SURVEY FINDINGS

Table K-1. Recommendations from Final Survey on Candidate Assessment Criteria

Initial Coding (Individual Comments)	Axial Coding (Dimensions)	Theoretical Coding (Factors/Categories)
1. Promotes balanced goals	Goal/objective setting	Strategic Leadership
2. Moderates the pace and scope of change	Change	Strategic Leadership
3. Communicates well	Communications	Communication (Blanchard & Associates, 2007; Drucker, 2008)*
4. Displays confidence	Confidence	Self-Mastery (Goleman, 1998)*
5. Promotes consensus	Consensus	Interpersonal Competence (Bass, 2008)
6. Maintains close contact with constituents	Contact with constituents	Communication
7. Displays courage in standing up to political pressure and protects people	Courage	Courage (Treasurer, 2008)*
8. Supports contributions to society in general	Contributions to society	Responsibility
9. Maintains credibility with the faculty	Credibility	Legitimization (Bass, 2008)* or Interpersonal Competence
10. Incorporates an academic model rather than a corporate chief executive officer model	Academics	University Mission Support
11. Displays cultural competence	Cultural competence	Interpersonal Competence
12. Delegates authority	Delegates authority	Strategic Leadership
13. Promotes the quality of education	Quality of education	Continuous Improvement
14. Maintains high ethical standards	Ethical standards	Consideration (Bass, 2008)
15. Honors the faculty role in shared governance	Respects shared governance	Strategic Leadership
16. Builds relationships with the faculty	Faculty relationships	Interpersonal Competence
17. Manages finances effectively	Financial management	Strategic Leadership or Stewardship (Gruder, 2008)
18. Promotes faculty diversity	Faculty diversity	Interpersonal Competence
19. Seeks participation in strategic planning	Strategic planning	Strategic Leadership
20. Inspires others	Inspires others	Interpersonal Competence
21. Focuses on institutional wellbeing and mission	Institutional wellbeing	University Mission Support
22. Listens	Listening	Communications
23. Does not micromanage	Delegates responsibility	Strategic Leadership
24. Promotes participative management	Participative management	Strategic Leadership
25. Keeps promises	Promise keeping	Consideration
26. Supports the Provost's role as the academic leader of the institution	Supports Leaders	Strategic Leadership
27. Serves the public	Public service	Social Responsibility
28. Represents the university	Represents university	University Mission Support
29. Emphasizes research as part of the institutional mission	Research emphasis	University Mission Support
30. Responds when issues brought up	Responsiveness	Interpersonal Competence
31. Serves as a role model	Role model	Consideration
32. "Thick-skinned"	Receptiveness to criticism	Interpersonal Competence
33. Provides thoughtful responses to issues	Thoughtful responses	Strategic Leadership
34. Promotes trust	Trust/trustworthiness	Interpersonal Competence
35. Demonstrates understanding	Empathy	Interpersonal Competence
36. Emphasizes education and research over athletics	University mission	University Mission Support
37. Successfully raises funds	Secures funding	University Mission Support
38. Makes himself/herself accessible to constituents	Accessibility	Communication
39. Demonstrates open mindedness	Open-mindedness	Consideration
40. Promotes transparency	Transparency/openness	Consideration
41. Recruits high quality faculty and staff	Recruiting	University Mission Support
42. Demonstrates understanding and respect for university policies, procedures, committee structures, and administrator areas of responsibility	Respect for roles and responsibilities	Stewardship
43. Promotes administrative efficiencies	Administrative effectiveness	Strategic Leadership

44. Demonstrates intelligence	Intelligence	Intelligence (Gardner, 2004; Goleman, 1998, 2006; Gruder, 2008)*
45. Follows good business practices	Business acumen	Stewardship
46. Displays honesty	Honesty	Consideration
47. Displays humanistic traits	Human relations	Interpersonal Competence
48. Display knowledge of the institutional culture	Cultural competence	University Mission Support
49. Gets involved in K-12 education	Promotes education in State	Responsibility

* Additional categories not identified in the eight-factor solution from PCA of quantitative data from the survey.

Table K-2. Recommendations from Final Survey Participants on External Factors

Recommended External Factor Considerations
1. Institutional type and size have a huge impact on the role of the president and how much they participate in daily operations such as recruiting students, faculty, and staff
2. Job complexities, economic drivers, and political drivers that limit the range of actions a president can take
3. Political pressures from state government
4. Agenda and pressures from the Board of Regents
5. Political influences and meddling

Table K-3. Recommendations from Final Survey Participants on Assessment Format

Recommended Assessment Format Characteristics
1. Assessment should involve multiple raters (e.g., 360 degree assessment)
2. Assessment should involve more than just members of the Board of Regents
3. Assessment process should involve focused groups
4. Assessment instrument should have quantifiable measures
5. Assessment instrument should be vetted by individuals with knowledge of survey instruments and avoid vagueness
6. Assessment instrument should include open-ended questions
7. Assessment should be a formal process
8. Assessment must take into account external factors such as the economy
9. Assessment instrument should include qualitative criteria
10. Assessment results (general) should be made public by the Board of Regents
11. Assessment results (specific) should remain private
12. Assessment should not include a form because it takes too many peoples' time
13. Assessments are best when they come from Regents trained in academics and involve individual faculty groups that are trained and aware of the issues
14. Assessment results must be reviewed with a critical eye and acted upon
15. Assessments should give greatest weight to inputs from students, faculty, and staff

Table K-4. Final Survey Comment Codes and Counts from Atlas *TI*[®]

Comment Code	Counts
Assessment criteria	37
Political influences	13
External factors	12
Survey questions	12
Assessment format	10
Athletics	10
Assessment participants	8
President responsibilities	8
Survey quality	7
University mission	7
Administrator compensation	6
Participative management	6
Shared governance	6
Accountability	5
President priorities	5
Academics	4
Communication	4
Education quality	4
Patronage appointments	4
360 degree assessment	3
Assessment results transparency	3
Faculty compensation	3
Financial management	3
Fundraising	3
Honesty	3
Open-ended questions	3
President compensation	3
Qualitative criteria	3
Research quality	3
Responsiveness	3
University focused	3
Accessibility	2
Assessment usefulness	2
Balanced institutional goals	2
Change	2
Courage	2
Ethical behavior	2
Goal and objective setting	2
Leadership characteristics	2
Listening	2
Open mindedness	2
President hiring	2
Provost roles and responsibilities	2
Quality of education	2
Quantitative criteria	2
Recruiting administrators	2
Scholarship	2
Student success	2
Teaching skills	2
Transparency	2
Understanding	2
Administrative efficiency	1

Comment Code	Counts
Administrator assessment	1
Administrator responsibilities	1
Administrator staffing	1
Advertising	1
Alignment	1
Assessment focused group	1
Autocratic leadership	1
Business CEO model	1
Confidence	1
Confidentiality	1
Consensus	1
Contact with constituents	1
Contributions to society	1
Credibility	1
Cultural competence	1
Decision-making	1
Delegates authority	1
Economy	1
Ethical standards	1
External role of president	1
Extracurricular activities	1
Faculty diversity	1
Faculty governance	1
Faculty quality	1
Faculty relationships	1
Focused groups	1
Formal assessment	1
Harmony	1
Hiring criteria	1
Humanistic traits	1
Inspiration	1
Institutional wellbeing	1
Institutional size	1
Intelligence	1
Involvement in K-12	1
Knowledge of institutional culture	1
Management style	1
Micromanagement	1
Morale	1
President image	1
President selection	1
President qualifications	1
Promise keeping	1
Public service	1
Recruiting faculty	1
Recruiting staff	1
Recruiting students	1
Representing the university	1
Role model	1
Service	1
Situational awareness	1
Soliciting opinions	1
Strategic planning	1
Thick-skinned	1
Thoughtful response	1

	Comment Code	Counts
Trust		1
Values		1
Works with others		1

APPENDIX L
FACTOR ANALYSIS FINDINGS

Table L-1. Factor Loading for Rotated Factor Structure from Responses to the Final Survey (N =280)

Rotated Component Matrix ^a								
Factor/Items	Component							
	1	2	3	4	5	6	7	8
Strategic Leadership								
ClearsObstacles	.704	.035	.103	.114	-.024	.070	.168	.074
AdaptsToChange	.676	.022	.106	.074	.273	.267	.024	.128
RealisticPlans	.662	-.079	.239	.317	.188	.087	-.125	.023
SetsGoalsObjectives	.645	.120	.255	.255	.060	.015	-.217	.104
Teambuilding	.609	.168	.036	.065	.202	.047	.075	.120
PerformanceIndicators	.589	.159	.324	.109	-.005	.094	.089	.082
SupportsOtherLeaders	.586	.317	.249	.096	.032	.050	.018	-.136
PromotesChange	.568	.072	.312	.061	.134	-.189	.184	-.017
PerformsRiskAnalysis	.541	.225	.209	.134	.138	.177	.081	.086
ResolvesConflicts	.538	.345	.038	.020	.153	.030	.040	.140
RewardsPerformance	.535	.241	.007	.189	-.038	.213	.285	-.240
ClearDirection	.513	.323	.048	-.023	.016	.363	.220	.066
ControlsBudget	.486	-.094	-.107	.390	.292	.207	-.102	.045
ThoughtfulResponses	.428	.372	.196	.016	.350	.172	.029	-.014
Consideration								
DisplaysCourtesy	.221	.724	.155	.040	.159	.174	.022	.030
SharesInformation	.094	.703	.144	.044	.190	.034	.130	.137
DisplaysIntegrity	.160	.692	-.060	.104	.087	.119	-.047	.121
Continuous Improvement								
UniversityStory	.295	-.049	.643	.203	.148	.125	-.169	-.080
PromotesEducation	.305	.164	.629	.118	.041	-.058	.200	.206
StrategicFramework	.462	.069	.533	.020	.245	.147	-.029	.115
LessonsLearned	.444	.289	.474	-.026	.103	.079	.054	.177
Benchmarking	.078	.083	.473	.342	-.033	.394	.223	.120
University Mission Support								
SecuresResources	.193	.002	.073	.673	-.077	.155	.093	.118
UniversityAdvocate	.035	.051	.308	.665	.104	.133	.116	-.085
RecruitsGoodPeople	.342	.207	-.021	.548	.158	-.081	.283	.000
JobCompetence	.269	.361	.105	.497	.032	-.151	.092	.297
Interpersonal Competence								
BuildsTrust	.231	.232	.186	.163	.687	.070	-.045	-.017

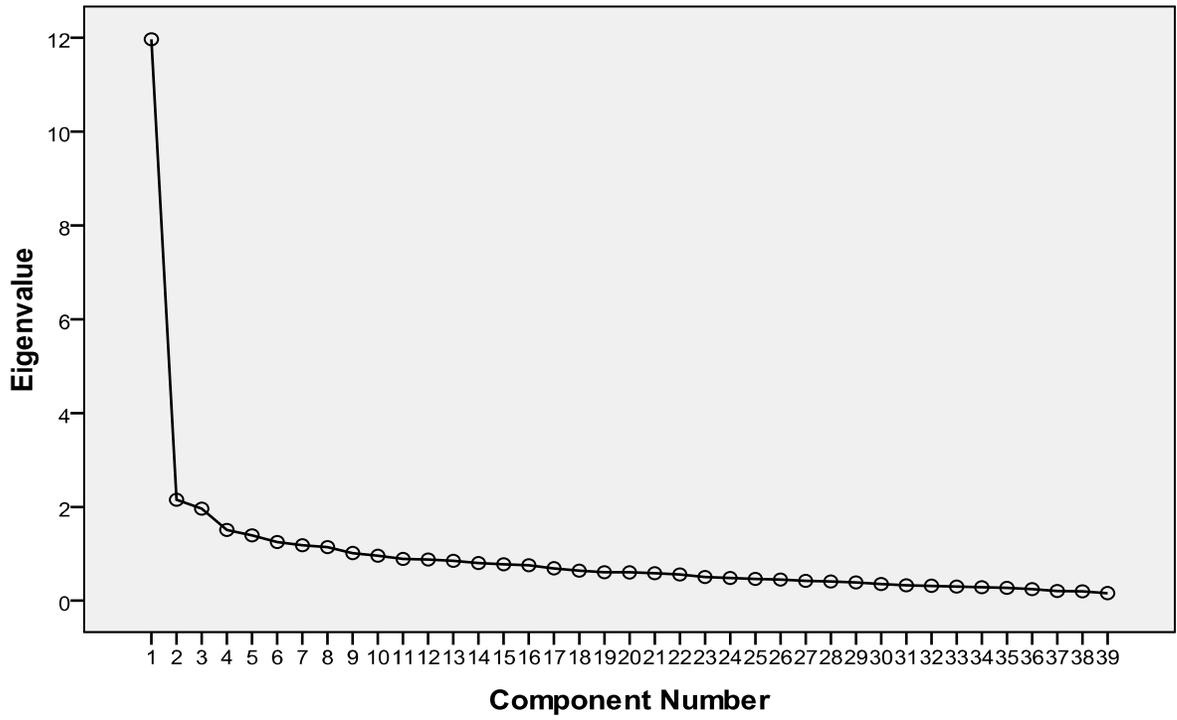
EmployeeDiversity	.212	.273	.008	-.030	.617	-.090	.212	-.011
Harmony	.097	.047	.130	-.009	.549	.298	.156	.279
Inspirational	.281	.264	.447	.116	.395	.120	.057	-.130
Stewardship								
ShowsAppreciation	.337	.263	.093	.069	.206	.624	.087	-.247
ResourceDecisions	.062	-.006	.131	.107	.023	.512	.138	.514
InformedDecisions	.380	.280	.030	.312	.076	.468	-.046	.126
ClimateAwareness	.132	.326	.149	.102	.123	.447	.147	.302
Academic Quality								
StudentFacultyRatio	.117	.109	-.212	.034	.156	.083	.707	.119
StudentQuality	.166	-.045	.301	.191	-.225	.063	.626	-.094
FacultyQuality	-.031	.029	.116	.135	.207	.077	.591	.120
Responsibility								
InstitutionalInterests	.046	.054	.037	-.007	.043	.016	-.013	.709
AdministratorSelection	.175	.296	.013	.125	.015	.032	.132	.607

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 12 iterations.

Table L-2. Scree Plot from Varimax Rotation of Final Survey Data



REFERENCES

- Agnes, M., & Guralnik, D. B. (Eds.). (2001). *Webster's new world college dictionary* (4th ed.). Foster City, CA: IDG Books Worldwide, Inc.
- Alfred, R. L. (2006). *Managing the big picture in colleges and universities: From tactics to strategy*. Westport, CT: American Council on Education/Praeger Publishers.
- Alfred, R., & Rosevear, S. (2000). Organizational structure, management, and leadership for the future. In A. M. Hoffman & R. W. Summers (Eds.), *Managing colleges and universities: Issues for leadership*, (pp. 1-28). Westport, CT: Bergin & Garvey.
- Alison, P. D. (2002). *Missing data: Series on quantitative applications in the social sciences*. Thousand Oaks, CA: SAGE Publications.
- American Association of University Professors. (2006). *Policy documents and reports*, (10th ed). Washington, DC: American Association of University Professors.
- American Educational Research Association (AERA), American Psychology Association (APA), & National Council on Measurement in Education (NCME). (2004). *Standards for educational and psychological testing*. Washington, DC: American Education Research Association.
- Arizona Board of Regents. (1990). *Evaluation of university presidents, executive director, and counsel to the board*. Retrieved February 21, 2008, from http://www.abor.asu.edu/1_the_regents/policymanual/Chap6/Chap6_part1.htm#6-106
- Arizona Board of Regents. (2005). *Multiple year employment contract for university president*. Retrieved September 25, 2009, from http://www.abor.asu.edu/1_the_regents/meeting/board_book/Jun-2007/Item-28-2007-06-Reappt-ASU-president-contract.pdf
- Armstrong, M. (2009). *Armstrong's handbook of performance management: An evidenced-based guide to delivering high performance*, (4th ed.). London: Kogan Page.
- Aron, A., & Aron, E. N. (2003). *Statistics for psychology* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Association of Governing Boards of Universities and Colleges (AGB). (2006). *The leadership imperative*. Washington, DC: AGB.

- Atwell, R. H. (2007, March/April). The craft of presidential assessment. *Trusteeship*, 28-32.
- Barker, J. F. (2006). A response to president Dubois' essay: Clemson and South Carolina's confederate flag. In D. G. Brown (Ed.), *University presidents as moral leaders* (pp. 59-62). Westport, CT: Praeger Publishers.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173-1182.
- Bass, B. M. (1990). *Handbook of leadership: Theory, research, & managerial application*, (3rd ed.). New York: The Free Press.
- Bass, B. M. (with Bass, R.) (2008). *The Bass handbook of leadership: Theory, research, & managerial applications* (4th ed.). New York: Free Press.
- Bellini, J. L., Rumrill, P. D., Jr., & Cook, B. G. (2001). Research Validity. In J. L. Bellini, P. D. Rumrill, Jr., & B. G. Cook, *Research validity in special education: Designs, methods, and applications*. Springfield, IL: Charles Thomas, Publishers.
- Berk, R. A. (Ed.). (1986). *Performance assessment: Methods and applications*. Baltimore, MD: The Johns Hopkins University Press.
- Bernardin, H. J. (1984). A performance appraisal system. In R. A. Berk, *Performance assessment*. Baltimore, MD: The Johns Hopkins University Press.
- Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent Publishing Company.
- Birnbaum, R. (1992a). *How academic leadership works: Understanding success and failure in the college presidency*. San Francisco: Jossey-Bass Publishers.
- Birnbaum, R. (1992b). Will you still love me in December as you did in May: Why experienced college presidents lose faculty support? *The Journal of Higher Education*, 63(1), 1-25.
- Blanchard, K. & Associates. (2007). *Leading at a higher level: Blanchard on leadership and creating high performance organizations*. Upper Saddle River, NJ: Blanchard Management Corporation.
- Borman, W. C. (1986). Behavior-based rating scales. In R. A. Berk (Ed.), *Performance assessment: Methods and applications*, pp. 100-120. Baltimore: The Johns Hopkins University Press.

- Bossidy, L., & Charan, R. (2002). *Execution: The discipline of getting things done*. New York: Crown Business.
- Boverie, P. E., & Kroth, M. (2001). *Transforming work: The five keys to achieving trust, commitment, and passion in the workplace*. Cambridge, MA: Perseus Publishing.
- Bowen, W. G., & Shapiro, H. T. (Eds.). (1998). *Universities and their leadership*. Princeton, NJ: Princeton University Press.
- Boynton, A., & Fischer, B. (2005). *Virtuoso teams: Lessons from teams that changed the world*. Harlow, England: Pearson Education Limited.
- Brint, S. (Ed.). 2002. *The Future of the City of Intellect: The Changing American University*. Stanford: Stanford University.
- Brown, D.G. (Ed.) 2006. *University Presidents as Moral Leaders*. Westport: Praeger Publishers.
- Brownell, J. (2005). Predicting leadership: The assessment center's extended role. *International Journal of Contemporary Hospitality Management*, 17(1), 7-21.
- Bruce, R. J. (2008). *Acting on promise: Reflections of a university president*. Philadelphia, PA: Polyglot Press, Inc.
- Bruns, W. J., Jr. (Ed.). (1992). *Performance measurement, evaluation, and incentives*. Boston: Harvard Business School Press.
- Butler, L. M. (2007). Warning lights: New dashboard reports help institutions gauge their performance. *Connections, Spring 2007*, 31-33.
- Byham, W. C., & Thornton G. C., III. (1986). Assessment centers. In R. A. Berk (Ed.), *Performance assessment: Methods and applications*, pp. 143-166. Baltimore: The Johns Hopkins University Press.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*, (2nd ed.). New York: Routledge.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Cannon-Bowers, J. A., & Salas, E. (2001). *Improving teamwork in organizations: Application of resource management training*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment: Series on quantitative applications in the social sciences*. Thousand Oaks, CA: Sage Publications, Inc.
- Center for Policy Analysis. (2007). *The American college president*. Washington, DC: American Council on Education.
- Cepeda, G., & Martin, D. (2005). A review of case studies publishing in Management Decision 2003-2004. *Management Decision*, 43(6), 851-876.
- Chaleff, I. (2003). *The courageous follower: Standing up to & for our leaders*, 2nd ed. San Francisco: Berrett-Koehler Publishers, Inc.
- Charmaz, K. (2004). Grounded theory. In Hesse-Biber & Leavy (Eds.), *Approaches to qualitative research: A reader on theory and practice* (pp. 496-521). New York: Oxford University Press, Inc.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: SAGE Publications, Inc.
- Charters, Jr., W. W. (1992). *Understanding variables and hypotheses in scientific research*. University of Oregon. (ERIC Document Reproduction Service No. ED342056)
- Clark, B. R. (2002). University transformation: Primary pathways to university autonomy and achievement. In S. Brint, S. (Ed.), *The future of the city of intellect: The changing American university*, (pp. 322-342). Stanford: Stanford University.
- Clemson University. (2009). Operating manual, Chapter X: Selection, duties, and evaluation of the president. Retrieved September 25, 2009, from <http://www.Clemson.edu/administration/bot/manual/chapter10.html>
- Coens, T. & Jenkins, M. (2000). *Abolishing performance appraisals: Why they backfire and what to do instead*. San Francisco: Berrett-Koehler Publishers, Inc.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd ed.). New York: Lawrence Erlbaum Associates, Publishers.
- Comrey, A. L. (1973). *A first course in factor analysis*. New York: Academic Press, Inc.
- Connolly, P. M., & Wilson, C. L. (2007). *Executive leadership: Forces for growth and globalization*. The Clark Wilson Group.
- Cooke, H. S., & Tate, K. (2005). *Project management: The McGraw-Hill 36-hour course*. New York: McGraw-Hill.

- Cranton, P. (2006). *Understanding and promoting transformational learning: A guide for educators of adults*, (2nd ed.). San Francisco: Jossey-Bass.
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks: SAGE Publications, Inc.
- Creswell, J. W., & Plano Clark, V.L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications, Inc.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity and psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- Crotty, M. J. (1998). *The foundation of social research: Meaning and perspective in the research process*. Thousand Oaks, CA: Sage Publications, Inc.
- Curnan, S., LaCava, L., Sharpsteen, D., Lelle, M., & Reece, M. (1998). *W.K. Kellogg Foundation Evaluation Handbook*. Battle Creek: W.K. Kellogg Foundation.
- Davis, W. E., & Davis, D. R. (1999). The university president: do evaluations make a difference? *Journal of Personnel Evaluation in Education*, 13(2), 199-140.
- Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.
- Dey, I. (1995). Reducing fragmentation in qualitative research. In U. Keele (Ed.), *Computer-aided qualitative analysis* (pp. 69-79). Thousand Oaks, CA: Sage.
- Dillman, D. D. (2007). *Mail and internet surveys: The tailored design method*, (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Driscoll, M. P. (2000). *Psychology of learning for instruction* (2nd ed.). Boston: Allyn and Bacon.
- Drucker, P. F. (2008). *Management*, (revised ed.). New York: HarperCollins.
- Duderstadt, J. J., & Womack, F. W. (2003). *The future of the public university in America*. Baltimore: The Johns Hopkins University Press.
- Engelkemeyer, S. W. (1999). Institutional performance measures. *American Association of Higher Education Bulletin*, 51(7), 3-6.
- Engelkeymer, S. W. (2008). Assessment workshop. Retrieved July 14, 2008, from <http://www.aacsb.edu/conferences/events/handouts/fac01/Handouts%20Engelkemeyer.pdf>

- Etzioni, A. (1964). *Modern organizations*. Englewood Cliffs, NJ: Prentice Hall.
- Fain, P. (2006, June). Relations between presidents and boards' top agenda at Chronicle leadership – 360-degree reviews are garbage, consultant says. *The Chronicle of Higher Education*, 2006(06). Retrieved August 25, 2006, from <http://chronicle.com/daily/2006/06/2006061601n.htm>
- Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw Hill.
- Field, A. (2005). *Discovering statistics* (2nd ed.). Thousand Oaks: SAGE Publications, Inc.
- Fine, S. A. (1986). Job analysis. In R.A. Berk (Ed.), *Performance assessment: Methods and applications*, (pp. 53-81). Baltimore: The Johns Hopkins University Press.
- Finkelstein, S., Hambrick, D. C., & Cannella, A. A. (2009). *Strategic leadership: Research on executives top management teams, and boards*. New York: Oxford University Press, Inc.
- Fisher, C. F., (Ed.). (1978). *New directions for higher education: Developing and evaluating administrative leadership*. San Francisco: Jossey-Bass Publishers.
- Fisher, J. L., Tack, M. W., & Wheeler, K. J. (1988). *The effective college president*. New York: Macmillan Publishing Company.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358.
- Florida State University (2004). *Employment agreement between the Florida State University Board of Trustees and Dr. Kent "T.K" Wetherell*. Lamont, FL: Florida State University.
- Fornier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions in Evaluation*, 68 (1995), 15-32.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.
- Gardner, H. (2004). *Frames of the mind: The theory of multiple intelligences*. New York: Basic Books.
- Gayle, J., Tewarie, B., & White, A. Q., Jr. (2003). Governance in the twenty-first-century university: Approaches to effective leadership and strategic management. *ERIC Digests*, ED482560, 1-6. Retrieved November 8, 2008, from ERIC database.

- George, D., & Mallery, P. (2007). *SPSS for Windows: Step by step: A simple guide and reference 14.0 update* (7th ed.). Boston: Pearson Education, Inc.
- Gibson, C. B. (1995). An investigation of gender differences in leadership across four countries. *Journal of International Business Studies*, 26(2), 255-279.
- Gini, A. (1995). Too much to say about something. *Business Ethics Quarterly*, 5(1), 143-155.
- Glaser, B. G., & Strauss, A. S. (1975). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.
- Goleman, D. (2007). *Social intelligence: The revolutionary new science of human relationships*. New York: Bantam Dell.
- Goleman, D. (2000). *Working with emotional intelligence*. New York: Bantam Books.
- Gorsuch, R. L. (1983). *Factor analysis*, (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Green, S. G., & Mitchell, T. R. (1979). Attributional processes of leaders and leader-member interactions. *Organizational Behavior and Performance*, 23, 429-458.
- Greenleaf, R. K. (2002). *Servant leadership: A journey into the nature of legitimate power and greatness*. New York: Paulist Press.
- Grote, D. (2002). *The performance appraisal: Question and answer book: A survival guide for managers*. New York: American Management Association.
- Gruder, D. (2008). *The new IQ*. Santa Rosa, CA: Elite Books.
- Guadagnoli, E., & Velicer, W. (1988). Relation of sample size to the stability of component patterns. *Psychology Bulletin*, 103(2), 265-275.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Hall, R. H. (1991). *Organizations: Structures, processes, and outcomes* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Hearn, T. K., Jr. (2006). Leadership and teaching in the American university. In D. G. Brown (Ed.), *University Presidents as Moral Leaders*, pp. 159-176. Westport: Praeger Publishers.

- Hearn, T. K., Jr., & Alexander, J. (2006). In D. G. Brown (Ed.), *University Presidents as Moral Leaders*, (pp. xi-xii). Westport: Praeger Publishers.
- Henning, G. (2007). Ordered response options. Retrieved October 17, 2009, from <http://www.dartmouth.edu/~saper/docs/Ordered%20Response%20Options.doc>
- Hesse-Biber, S. N., & Leavy, P. L. (2006). *The practice of qualitative research*. Thousand Oaks, CA: Sage Publications, Inc.
- Hoffman, A. M, & Summers, R. W. (Eds.). (2000). *Managing Colleges and Universities: Issues for Leadership*. Westport: Bergin & Garvey.
- Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *American Psychologist*, *June 1994*, 1-33.
- Honeyman, D. S. (2000). The financing of higher education. In A. M. Hoffman & R.W. Summers (Eds.), *Managing Colleges and Universities: Issues for Leadership*, (pp. 89-102). Westport: Bergin & Garvey.
- Illgen, D. R., & Favero, J. L. (1985). Limits in generalization from psychological research to performance appraisal processes. *The Academy of Management Review*, *10*(2), pp. 311-321.
- Ingram, R. T., & Weary, W. A. (with McPherson, M.P, & Munitz, B.) (2000). *Presidential and board assessment in higher education: Purposes, polices, and strategies*. Washington, DC: Association of Governing Boards of Universities and Colleges.
- Jacobs, R. R. (1986). Numerical rating scales. In R. A. Berk (Ed.), *Performance assessment: Methods and applications*, (pp. 82-89). Baltimore: The Johns Hopkins University Press.
- Johnson R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.
- June, A.W. (2007, February 16). Presidents: Same look, different decade. *The Chronicle of Higher Education*, p. A33-A35.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 32-36.
- Kaiser, R. B., Hogan, R., & Craig, S. B. (2008). Leadership and the fate of organizations. *American Psychologist*, *63*(2), 96-110.

- Kass, R. A., & Tinsley, E. A. (1979). Factor analysis. *Journal of Leisure Research, 11*(2), 120-138.
- Kaufman, J. F. (1978). Presidential assessment development. In C. F. Fisher (Ed.), *New directions in higher education: Developing and evaluating administrative leadership*, (pp. 59-69). San Francisco: Jossey-Bass Publishers.
- Kelle, E. (Ed.). (1995). *Computer-aided qualitative data analysis*. Thousand Oaks, CA: Sage.
- Kellerman, B. (2008). *Followership: How followers are creating change and changing leaders*. Boston: Harvard Business Press.
- Kent, T. W., Crotts, J. C., & Azziz, A. (2001). Four factors of transformational leadership behavior. *Leadership and Organization Development Journal, 22*(5), 221-229.
- Keohane, N. O. (2006). *Higher ground: Ethics and leadership in the modern university*. Durham, NC: Duke University Press.
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook*, (4th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Kerr, C. (1984). *Presidents make a difference: Strengthening leadership in colleges and universities, a report of the commission on strengthening presidential leadership*. Washington, DC: Association of Governing Boards of Universities and Colleges.
- Kerzner, H. (2006). *Project management: A systems approach to planning, scheduling, and controlling*, (9th ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Khurana, R. (2003). The curse of the superstar CEO. In *Harvard Business Review on Leadership at the Top*, pp. 1-18. Boston: Harvard Business School Publishing.
- Klenke, K. (2008). *Qualitative research in the study of leadership*. Bingley, UK: Emerald Group Publishing Limited.
- Kline, P. (1999). *The handbook of psychological testing*, (2nd ed.). London: Routledge.
- Knowles, M.S., Holton, E.F. III, & Swanson, R.A. (2005). *The adult learner: A definitive classic in adult education and human resource development*, (6th ed.). Burlington, MA: Elsevier.
- Kotter, J.P. (1988). *The leadership factor*. New York: The Free Press.

- Kotter, J. P. (1990). *A force for change: How leadership differs from management*. New York: The Free Press.
- Kotter, J. P. (1998). What leaders really do. In *Harvard Business Review on Leadership*, pp. 37-60. Boston: Harvard Business School Publishing.
- Kouzes, J. M., & Posner, B. Z. (2007). *The leadership challenge*, (4th ed.). San Francisco: Jossey-Bass.
- Krisch, J. (2008). Trustees OK presidential evaluation policy. Retrieved February 21, 2008, from <http://www.ohio.edu/outlook/07-08/February/320.cfm>
- Kuhn, T. S. (1996). *The structure of scientific revolutions*, (3rd ed.). Chicago: The University of Chicago Press.
- Latham, G. L., & Wexley, K. N. (1981). *Increasing productivity through performance appraisal* (2nd ed.). Reading, MA: Addison-Wesley Publishing Company.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Learning Center (2008). Leadership assessment: Personal satisfaction survey. Retrieved March 31, 2008, from <http://www.learning.center.net/library/leadership.shtml>
- Lecky-Thompson, R. (1999). *Constructive appraisals: The essential guide to thinking and working smarter*. New York: Marshall Editions Development Ltd.
- Lencioni, P. (2002). *The five dysfunctions of a team: A leadership fable*. San Francisco: Jossey-Bass.
- Leske, J. S. (1991). Internal psychometric properties of the critical care family needs inventory. *Heart and Lung*, 20(3), 236-244.
- Levin, J. S. (2000). The practitioner's dilemma: Understanding and managing change in academic institutions. In A.M. Hoffman & R.W. Summers (Eds.), *Managing Colleges and Universities: Issues for Leadership*, (pp. 29-42). Westport: Bergin & Garvey.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills: SAGE Publications, Inc.
- Locher, T. K., Jr., & Alexander, J. (2006). Assessment: Appraisal trends. *Personnel Journal*, 60, 10-13.

- Longnecker, C. O. (1997). Why managerial performance appraisals are ineffective: Causes and lessons. *Career Development International*, 2(5), 212-218.
- MacCallum, R. C., Widaman, K. F., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99.
- Mason, J. (2006). Mixing methods in a qualitatively drive way. *Qualitative Research*, 6(1), 9-25.
- McNeese State University. (2005). Policies and procedures for the evaluation of the president and provost. Retrieved September 25, 2009, from <http://www.mcneese.edu/policy/docs/PresProvEval.pdf>
- Messick, S. (1980). Test validity and ethics assessment. *American Psychologist*, 35(11), 1012-1027.
- Merriam, S. B., & Caffarella, R. S. (1999). *Learning in adulthood: A comprehensive guide*, (2nd ed.). San Francisco: Jossey-Bass.
- Mezirow, J., & Associates. (2000). *Learning as transformation: Critical perspectives on a theory in progress*. San Francisco: Jossey-Bass.
- Michael, S. O., Schwartz, M., & Balraj, L. (2001). Indicators of presidential effectiveness: a study of trustees of higher education institutions. Retrieved February 15, 2006, from <http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/0600150702.htm>
- Miller, R., & Miller, P. (2000). Evaluating collegiate administrators. In A. M. Hoffman, & R.W. Summers, (Eds.), *Managing colleges and universities: Issues for leadership*, (pp. 199-210). Westport, CT: Bergin & Garvey.
- Mintzberg, H. (1998). *The manager's job: Folklore and fact*. In *Harvard Business Review of Leadership*, (pp. 1-36). Boston: Harvard Business School Publishing.
- Moberg, D. J. (1999). The big five and organizational virtue. *Business Ethics Quarterly*, 9(2), 245-272.
- Mohrman, A. M., Jr., Resnick-West, S. M., & Lawler, E. E., III. (1989). *Designing performance appraisal systems: Aligning appraisals and organizational realities*. San Francisco: Jossey-Bass Publishers.
- Munitz, B. (1978). Strengthening institutional leadership. In C. F. Fisher (Ed.), *New directions in higher education: Developing and evaluating administrative leadership*, (pp. 11-21). San Francisco: Jossey-Bass Publishers.

- Nason, J. W. (1997). *Presidential assessment: A guide to periodic review of the performance of chief executives*. Washington, DC: Association of Governing Boards of Universities and Colleges.
- Nathan, B. R., & Cascio, W. F. (1986). Introduction: Technical and legal standards. In R. A. Berk (Ed.), *Performance assessment: Methods and applications*, (pp. 1-50). Baltimore: The Johns Hopkins University Press.
- New Mexico Attorney General. (2009). Open meeting act and inspection of public records act (NM sunshine laws). Retrieved September 25, 2009, from http://www.nmag.gov/office/Divisions/Civ/OMA_IPRA/default.aspx
- New Mexico State University. (2009). Quick facts 2008-2009. Retrieved August 22, 2009 from http://www.nmsu.edu/Research/iresearch/QuickFacts/Factsheets/Fact_Brochure_Fall_08-12-24-08.pdf
- Northouse, P. G. (2007). *Leadership: Theory and practices*, (4th ed.). Thousand Oaks, CA: SAGE Publications.
- Nunnally, J. (1978). *Psychometric theory*, (2nd ed.). New York: McGraw-Hill Publishing Company.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*, (3rd ed.). New York: McGraw-Hill.
- Ohio University. (2006). President McDavis receives positive evaluation from the university's board of trustees. Retrieved June 17, 2008, from <http://www.ohio.edu/outlook/06-06/June/573n-056.cfm>
- Oppenheim, A. N. (1966). *Questionnaire design and attitude measurement*. New York: Basic Books, Inc.
- Ory, J. C. (2000). Managing evaluations in higher education. In A. M. Hoffman & R. W. Summers (Eds.), *Managing colleges and universities: Issues for leadership*, (pp. 187-198). Westport, CT: Bergin & Garvey.
- Padilla, A. (2005). *Portraits in leadership: Six extraordinary university presidents*. Westport, CT: Praeger Publishers.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*, (3rd ed.). USA: Thompson Learning.

- Pett, M. A, Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: SAGE Publications.
- Pettijohn, L. S., Parker, R. S., Pettijohn, C. E., & Kent, J. L. (2001). Performance appraisals: Usage, criteria, and observations. *Journal of Management Development, 20*(9), 754-771.
- Random.Org (2009). Integer generator. Retrieved June 27, 2009 from <http://Random.Org>
- Rea, L. M., & Parker, R. A. (2005). *Designing and conducting survey research: A comprehensive guide* (3rd ed.). San Francisco: Jossey-Bass.
- Rhodes, F. T. (2001). *The creation of the future: The role of the American university*. Ithaca: Cornell University Press.
- Rost, J. C. (1993). Leadership development in the new millennium. In B. M. Bass & R. Bass, *The Bass Handbook of Leadership*, p. 15. New York: Free Press.
- Rudestam, K. E., & Newton, R. R. (2001). *Surviving your dissertation: A comprehensive guide to content and process*, (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Russ-Eft, D., & Preskill, H. (2001). *Evaluation in organizations: a systematic approach to enhancing learning, performance, and change*. New York: Basic Books.
- Sahakian, W. S., & Sahakian, M. L. (2005). *Ideas of the great philosophers*. New York: Fall River Press.
- Salazar, M. (2009, June 3). Regents give Schmidly positive evaluation. *The Albuquerque Journal*, pp. C1-C2.
- Sanaghan, P. H., Goldstein, L., & Gavel, K. D. (2008). *Presidential transitions: It's not just the position, it's the transition*. Westport, CT: Praeger Publishers.
- Schwartz, M. P. (1998). Assessing the performance of academic presidents. *Dissertation Abstracts International*. (UMI No. 9921004).
- Schwartz, M. P. (2001). *Annual presidential performance reviews*. Washington, DC: Association of Governing Boards of Universities and Colleges.
- Scriven, M. (1980). *The logic of evaluation*. Point Reyes, CA: Edgepress.

- Scriven, M. (1993). Hard-won lessons in program evaluation. *New Directions for Program Evaluation*, 58(summer 1993), 1-107.
- Scriven, M. (1995). The logic of the evaluation process. *New Directions for Program Evaluation*, 68(1995), 49-70.
- Scriven, M. (2007). The logic of evaluation. Retrieved October 30, 2008, from <http://homepages.wmich.edu/~mscriven/THE%20LOGIC%20OF%20EVALUATIONS.doc>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, 19(1993), 405-450.
- Siegle, D. (2009). Likert scale. Retrieved October 17, 2009, from <http://www.gifted.uconn.edu/Siegle/research/Instrument%20Reliability%20and%20Validity/Likert.html>
- Slife, B. D., & Williams, R. N. (1995). *What's behind research? Discovering hidden assumptions in qualitative analysis*. Thousand Oaks, CA: Sage.
- Sokol, M., & Oresick, R. (1986). Managerial performance appraisal. In R. A. Berk (Ed.), *Performance assessment: Methods and applications*, (pp. 376-392). Baltimore: The Johns Hopkins University Press.
- Southeastern Louisiana University. (1997). Policy and procedures for the evaluation of the president and vice presidents. Retrieved March 6, 2006, from <http://www2.selu.edu/Administration/Inst-Research/IE/admneval.htm>
- SPSS, Inc. (2010). *PASW[®] missing values 18*. Chicago: SPSS, Inc.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park: SAGE Publications, Inc.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- StudentVoice[™]. (2008). *Technical information*. Buffalo, NY: StudentVoice[™].

- Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education, Inc.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology*. Thousand Oaks: SAGE Publications, Inc.
- Thorndike, R.M. (2005). *Measurement and evaluation in psychology and education* (7th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Tinsley, H. E., & Tinsley, D. T. (1987). Using factor analysis in consulting psychology research. *Journal of Counseling Psychology*, 34(4), 414-424.
- Trachtenberg, S. J. (2008). *Big man on campus: A university president speaks out on higher education*. New York: Touchtone.
- Treasurer, B. (2008). *Courage goes to work*. San Francisco: Berrett-Koehler Publishers, Inc.
- Trow, M. (1998). On accountability of higher education in the United States. In W. G. Bowen & H. T. Shapiro (Eds.), *Universities and their leadership*, (pp. 15-64). Princeton, NJ: Princeton University Press.
- Uniform guidelines on employee selection procedures. *Federal Register*, August 25, 1978, 43, 38290-38315.
- University of Alabama System. (2003). Board of trustees manual. Retrieved February 24, 2006, from <http://sacs.ua.edu/uasystem/boardoftrustees/manual/articlev.html>
- University of Michigan. (2006). Evaluation: President evaluation form. Retrieved July 7, 2006, from http://aec.umich.edu/leo/president_v02.html
- University of Nevada Las Vegas. (2000). Faculty evaluation of the president. Retrieved March 6, 2006, from <http://facultysenate.unlv.edu/evaluation.html>
- University of Nevada Las Vegas. (2003). Faculty evaluation of the president. Retrieved March 6, 2006, from <http://facultysenate.unlv.edu/preeval2003.htm>
- University of New Mexico (2002). University business policies and procedures manual: performance review and recognition. Retrieved February 23, 2006, from <http://www.unm.edu/~ubppm>
- University of New Mexico. (2004). Board of regents policy manual, 1.1. Retrieved February 23, 2006, from <http://www.unm.edu/~brpm/r11.htm>

- University of New Mexico (2007a). Five-year academic ledger: Albuquerque campus. Retrieved April 12, 2008, from <http://www.unm.edu/documents/>
- University of New Mexico (2007b). Regent's goals and evaluation criteria for the president of the University of New Mexico. Retrieved, April 12, 2008, from <http://www.unm.edu/documents/>
- University of New Mexico (2008). Key dashboard indicators of progress toward UNM presidential work plan: 2007/2008 review AY 2007-2011. Retrieved October 4, 2008, from <http://www.unm.edu/president/documents/>
- University of New Mexico (2009). President's work plan for FY10. Retrieved October 3, 2009, from <http://www.unm.edu/president/documents/>
- University of New Mexico (2010). Five-year academic ledger: Albuquerque campus. Retrieved March 30, 2010, from <http://www.unm.edu/president/documents/>
- University of North Florida (2006). Presidential evaluation task force: Report and recommendations. Retrieved July 7, 2006, from <http://www.unf.edu/trustees/presevalplan/presevalplan.pdf>
- University of Utah System Board of Regents. (2005). Evaluation of presidents. Retrieved February 21, 2008, from <http://www.utahsbr.edu/policy/r209.htm>
- University of Washington Board of Regents. (2006). 2005-2006 President evaluation. Retrieved September 25, 2009, from <http://www.washington.edu/regents/meetings/meetings06/July7items/c-1.pdf>
- U. S. News and World Report (2008). Undergraduate ranking criteria and weights. Retrieved January 19, 2008, http://colleges.usnews.rankingsandreviews.com/usnews/edu/college/rankings/about/weight_print.htm
- van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. Ontario, Canada: The University of Western Ontario.
- VisionMetrics. (2008). HR solutions survey. Retrieved March 31, 2008, from <http://www.visionmetrics.net/en/hr-survey-solutions.aspx>
- Wagner, S. F. (1992). *Introduction to statistics*. New York: HarperCollins Publishers, Inc.
- Weiss, D. J. (1971). Research methodology: Further considerations in applications of factor analysis. *Journal of Counseling Psychology*, 18(1), 85-92.

- Wexley, K. N. (1986). Appraisal interview. In R. A. Berk (Ed.), *Performance assessment: Methods and applications* (pp. 167-202). Baltimore, MD: The Johns Hopkins University Press.
- Wilkinson, L., & Task Force on Statistical Inference (1999). Statistical methods in psychological journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604).
- Yeh, Q. L. (1995). Leadership, personal traits, and job characteristics in R&D organizations: A Taiwanese case. *Leadership and Organizational Journal*, 16(6), 16-26.
- Yu, H., & Miller, P. (2005). Leadership style: The X generation and baby boomers compared in different cultural contexts. *Leadership and Organizational Development Journal*, 26(1), 35-50.
- Yuchtman, E., & Seashore, S. E. (1967). A system resource approach to organizational effectiveness. *American Sociological Review*, 32(6), 891-903.
- Yudof, M. G., & Busch-Vishniac, I. J. (1996). Myth or management in universities? *Change*, 28(6), 18-27.
- Zaleznik, A. (1998). Managers and leaders: Are they different? In *Harvard Business Review on Leadership*, (pp. 61-88). Boston: Harvard Business School Publishing.
- Zemsky, R., Wegner, G. R., & Massey, W. F. (2005). *Remaking the American university: Market-smart and mission-centered*. New Brunswick, NJ: Rutgers University Press.