

To the Graduate Council:

I am submitting herewith a dissertation written by Kristi Jolena Nelms entitled "A Reliability and Validation Study of an Administrative Effectiveness Survey for Deans at Institutions of Higher Education." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Educational Administration and Policy Studies.



E. Grady Bogue, Major Professor

We have read this dissertation and
recommend its acceptance:

Norma J. Mertz
Walt Brant

Marcy Sue Greenger

Accepted for the Council:



Vice Chancellor and Dean of
Graduate Studies

**A RELIABILITY AND VALIDATION STUDY OF AN ADMINISTRATIVE
EFFECTIVENESS SURVEY FOR DEANS AT INSTITUTIONS OF HIGHER
EDUCATION**

A Dissertation
Presented for the
Doctor of Education
Degree
The University of Tennessee, Knoxville

Kristi Jolena Nelms
August 2005

Copyright © by Kristi Jolena Nelms
All rights reserved.

Thesis
2005b
.N44

DEDICATION

This dissertation is dedicated to my parents,

Charles Nelms and Betty Nelms,

who have provided me with the love, support, and knowledge to know that I can achieve
my dreams.

ACKNOWLEDGEMENTS

The support of many people helped make this dissertation a reality. First, I would like to express my sincere appreciation to my chairperson, Dr. E. Grady Bogue, who helped me to achieve this goal. I want to thank him for challenging me as well as supporting me in this endeavor. Most important, I want to thank him for believing in me and never giving up on me.

I also want to thank my other committee members: Dr. Norma Mertz, Dr. Mary Sue Younger, and Dr. Bill Lyons. They provided me with much needed assistance, patience, and encouragement.

Next, I am very grateful to my family, who always provided an encouraging word when I wanted to give up. These persons include my brother Chuck, my brother Terry and his family, Lynne, Lauren, and Ashlee, my aunt Francis, my uncle Billy, my aunt Betty Ann and my uncle Carrel.

Special thanks is extended to my closest personal friends Cole and Anna Lisa Stanley, and their children, Caleb and Audrey, who kept me focused and helped me to laugh when I wanted to cry. Thank you all for your endless expressions of care, support, and love.

Other special friends who deserve recognition for their reassurance and encouragement include Lesley and Kevin Cole, Robert Powell, Mark Roy, Phyllis and Alan Culvahouse, and Claire Combs. They always provided a listening ear when needed.

Long distance friends like Krista and Matt Smith, Thelma Clayton, Kathy and Kent Sherman, Stacey and Charlie Mauch, Darin Sherman, and Genevieve Goodman also deserve mention for their needed words of reassurance and love.

Additionally, much appreciation is extended to Kathy Morgan, who always encouraged me and helped get me through some difficult times. Through the years, there were many others in the Department of University Housing at the University of Tennessee who always provided their support.

Appreciation is also extended to my employer, Mr. Jim Grubb, Assistant Vice Chancellor for Student Affairs and Executive Director of University Housing at the University of Tennessee, Knoxville, for his support and cooperation.

ABSTRACT

The purpose of this study was to revise and to establish the reliability and validity of the *Administrative Effectiveness Survey* and to field test the survey on the performance of academic deans at selected institutions of higher education. Content validity for the instrument was, in part, established by asking two expert panels to review the instrument and make comments. Revisions to the initial instrument were made based on the suggestions by those experts. As a further test of content validity, factor analysis was performed on the data from the national study.

The instrument was administered to 677 faculty members at four colleges and universities across the country and 179 faculty members completed the instrument for a response rate of 26%. The instrument contained five different factor areas, planning and goal setting, leadership style and skill, resource stewardship, values and standards, academic stewardship, and a question about the overall performance of the academic dean. Results from the factor analysis revealed one primary factor, leadership, and two much smaller and less significant factors, external academics and mission, standards, and resources. The data were also subjected to reliability testing. A reliability coefficient of .983 was produced by the Cronbach Alpha.

Low response rates by the national expert panel, institutional participation, and respondents to the instrument resulted in less than desired outcomes of the study. The statistical analyses, while meeting minimal qualifications for conducting them, were not strong indicators of validity and reliability. Thus, the *Administrative Effectiveness Survey* cannot be deemed valid and reliable at this time without further research.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
	Statement of Problem.....	7
	Purpose.....	10
	Research Questions	10
	Significance of the Study	11
	Assumptions.....	11
	Limitations and Delimitations.....	12
	Definitions of Terms	13
II.	REVIEW OF THE LITERATURE.....	15
	Accountability.....	15
	Personnel Evaluation	21
	Faculty.....	21
	Administrators.....	23
	Academic Deans	27
	Evaluation Instruments	35
	Conclusion	40
III.	RESEARCH DESIGN AND METHOD	41
	Introduction.....	41
	Instrument Revision.....	42
	Instrument Validity	44
	Instrument Reliability	47
	Instrument Pilot Study	48
	Site and Population	50
	Data Collection Procedures.....	51
	Treatment of the Data	53
IV.	ANALYSIS OF THE DATA.....	56
	Introduction.....	56
	Research Question #1	56
	Research Question #2	59
	Research Question #2a.....	65
	Research Question #2b	67
	Research Question #3	70
	Summary	80
V.	SUMMARY, FINDINGS, DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS.....	85
	Summary	85
	Findings.....	87

Discussion	89
Conclusions.....	94
Recommendations.....	95
List of References	97
Appendices.....	108
Appendix A – Faculty Senate Administrative Questionnaire.....	109
Appendix B –	
<i>ADMINISTRATIVE EFFECTIVENESS SURVEY – ONE</i>	
<i>ADMINISTRATIVE EFFECTIVENESS SURVEY – TWO.....</i>	<i>112</i>
Appendix C – Letter to Members of the Expert Panels	119
Appendix D – Letter to Initial Review Panel.....	121
Appendix E – Letter to Participating Institutions	123
Appendix F – Institutional Consent to Participate Form	125
Appendix G – E-mail to Faculty Members.....	127
Appendix H – Cover Letter to Faculty Members	
(Informed Consent Letter)	129
Appendix I – Follow-up E-mail to Faculty.....	132
Appendix J – Final Version of the <i>Administrative Effectiveness</i>	
<i>Survey.....</i>	<i>134</i>
Appendix K – Revised <i>Administrative Effectiveness Survey</i>	
with Three Factors	138
Appendix L – Summary of Demographic Information for	
the National Study.....	141
Appendix M – Suggestions for Additional Questions to be	
Added to the Instrument.....	144
Appendix N – Recommendations to Improve the Instrument	147
Vita	150

LIST OF TABLES

Table	Page
1. Pearson Correlations – Pilot Study	60
2. Principal Component Analysis	62
3. Rotated Component Matrix.....	64
4. Multivariate Tests – Pilot Study	66
5. Multiple Regression – Pilot Study	68
6. Multiple Regression – National Study, Original Data.....	69
7. Multiple Regression – National Study, Three New Factors	71
8. Reliability Statistics – Pilot Study	72
9. Reliability Statistics for Subcategories for Pilot Study.....	73
10. Reliability Statistics – National Study.....	76
11. Reliability Statistics for Subcategories for National Study	77
12. Reliability Statistics for New Factors for National Study	81
13. Responses Regarding Two Open-Ended Questions	84

CHAPTER I

INTRODUCTION

The emergence of formal administrator performance appraisal occurred primarily in the latter half of the 20th century as a part of the accountability movement (Wattenbarger, 1983). Accountability policies in higher education emerged as a result of several factors: competition for resources from such areas as health care and public safety, reduction of financial support from both the state and federal governments, and unethical acts on the part of both faculty members and administrators (Ewell, 1994; Bogue & Noland, 2003). In addition, a demand for accountability gained momentum in the 1990s when some observers noted that society, rather than institutions of higher education, was the primary initiator of change due to three reasons: 1) society does not have many new resources to distribute; 2) society has more consumers of those resources; and 3) American society demands that higher education support the economy due to intensified international economic competition (Kerr, 1994). Institutions of higher education have an organizational responsibility to define their mission and purpose (what are they trying to achieve?) as well as to assess and evaluate their progress (how good a job are they doing?). Accountability issues, including programmatic and personnel issues, should be addressed (Bogue & Hall, 2003). Performance appraisal of university administrators is one area of accountability that can and should be a focus in higher education.

While much of what has been written on the emergence of accountability in American higher education has centered on access to colleges and universities and learning outcomes of college students, evaluation interest has also centered on faculty

and administrator performance. Historically, the evaluation of faculty members was more prevalent and part of the institutional culture prior to the advent of administrator evaluations. By the 1980s, faculty evaluation had become institutionalized primarily because of tenure and promotion requirements (Stronge & Helm, 1991). Students commonly use rating scales to evaluate faculty performance, and there are a number of validated and standardized forms for this purpose (Wattenbarger, 1983). Although there has been and continues to be a lively debate over the extent, for example, to which students are informed evaluators of faculty performance (Cahn, 1994), a range of institutional and commercial instruments for student assessment of faculty performance and instruction has emerged.

The argument of “what’s good for the goose is good for the gander” did not take long to emerge. If administrators were to evaluate faculty, then surely turn about would become fair play. Faculty members wanted the opportunity to provide feedback on the performance of administrators and university administrators wanted the chance to garner feedback to assist them with improvement of their performance. Over the last 30 years, colleges and universities have focused on developing formal administrator evaluation systems because of an increasing internal demand for accountability by both faculty and administrators. The increasing application of business practices to higher education institutions may have also contributed to some of this focus on formal administrator evaluation systems (Lahti, 1978; McKinnon, 1993).

With respect to administrative performance evaluation instruments, Wattenbarger (1983) commented 20 years ago that few validated or standardized scales existed and not much has changed since then. One of the most frequently used evaluative instruments is

the evaluation survey in which administrators' constituents are asked to furnish evaluative feedback on a number of factors deemed relevant to administrator effectiveness. For example, Budig (1995) described the system adopted at Vincennes University that allowed all faculty and professional staff to evaluate each administrator above them. The evaluation method used in their system was a single instrument containing 22 statements that evaluators could use to rate the performance of administrators' managerial responsibilities. Examples of the statements included: supportive of the college mission, meets objectives of position, creates an environment to foster new approaches, encourages participation, and maintains high standards of ethics. Seventy-five percent of the respondents selected strongly agree to the statements and 25% agreed to the statements. The question of whether such evaluative instruments are valid or reliable remains more elusive and Budig did not conduct any reliability or validity studies of this system.

Barrett (1993) provided several reasons for the need for administrator evaluation. One reason was a greater government regulation of personnel functions, in particular the move toward equal employment opportunity and affirmative action. Another reason was the trend toward greater employee participation in organizational decision making. One other reason was the perception that improved personnel practices can improve productivity. Mohrman, Resnick-West, and Lawler (1989) also listed explanations for conducting performance appraisal: a well-done appraisal has the potential to increase both individual and organizational performance, appraisals are designed to facilitate information exchange and influence performance, and appraisals can set expectations for employee performance. However, while such rationales exist for the use of administrator

performance evaluation, the evaluation instruments utilized at most institutions are not validated by research (Ory, 2000). The purpose of an administrator evaluation is to deliver:

a useful and trustworthy evaluation to the administrator to whom the person being evaluated reports, to provide information to administrators with the intent of helping them better understand their competencies as leaders, and to promote a more productive working relationship with the unit faculty or staff and the administrator being evaluated (Ory, 2000, p. 189).

Ory (2000) claimed that the successful use of evaluation instruments should not be confused with controlled, statistical research.

Many institutions have some type of administrator evaluation system in place (Budig, 1995; Belcher, 1992; Lahti, 1978; Brown & Rodriguez, 1989; Ory, 2000; Wilson & Kroesser, 1981). For example, in the early 1990s at the University of Tennessee, a subcommittee of the Faculty Senate developed an instrument for the evaluation of administrator performance, but no validity or reliability studies were ever completed for that instrument. A study by Brown and Rodriguez (1989) described the creation of a self-appraisal instrument and established its reliability and validity; however, this is an exception rather than the norm. Most institutional administrative evaluation systems do not have established reliability or validity (Ory, 2000).

The academic dean is the least studied of all the major academic positions and is the focus of this study (Bowker, 1982). An academic dean is in a unique position: serving as the executive head of a college or school; holding responsibility as an academic statesman including managing instructional services and fundraising; participating in conversations of resource allocations and competing for resources for his or her college; bridging the interests between faculty members and other university

groups; and serving as a member of the president's staff. The dean's role has shifted in many cases from "the senior scholar in a school to a senior manager in the university" (Austin, Ahearn, & English, 1997, p. 1). Demands such as fundraising, personnel decision-making, recruiting diverse faculty members, representing the university community, and even pressure to develop relationships with external constituencies to find potential donors have been added to the dean's job (Sarros, Gmelch, & Tanewski, 1998; Wolverton, Montez, & Gmelch, 2000). At the executive level of the dean, interpersonal and managerial skills are preeminent in creating a climate in which department chairs and faculty members can work effectively. In fact, Birnbaum (1992) conducted a study of academic leadership, and interviewees were asked to name important leaders on their campuses. Forty-one percent named the dean; and in large, decentralized institutions, deans were mentioned more frequently than anyone else, including the president. What little research that has been conducted with regard to the position of the academic dean has focused on the roles and traits of deans (Martin, 1993).

In the mid-Seventies, the systematic evaluation of academic deans began to receive some attention by researchers and university officials, primarily because of the continued focus on faculty evaluation (Rasmussen, 1978). During this time, only about 17% of institutions formally evaluated academic deans (Surwill & Heywood, 1976). By the 1980s, the evaluation of administrators had expanded and only 25% to 28% of academic deans were not formally evaluated (Bowker, 1982; Lynch & Bowke, 1984). The President or the Chancellor of the institutions evaluated the majority of academic deans and, in one-third of the institutions, the Vice President or the Provost conducted the evaluations of the dean (Bowker, 1982). Rasmussen (1978) quoted one dean: "As an

academic dean, one of the worst features of my job is that I never know how well I am doing it....Now, as a dean, I find that people do not tell me how they feel about my administrative style or the quality of my work” (p. 23). Many college presidents provide support and encouragement to their academic deans rather than provide the constructive criticism that some deans would prefer (Rasmussen, 1978). Some institutions do not make explicit any formal criteria or procedures for evaluating a dean’s performance (Lasley & Haberman, 1987).

Most institutions do have some kind of evaluation process in place for the evaluation of their academic deans; however, these processes are varied and rarely validated by research (Bowker, 1982). The University of Tennessee, Knoxville has such a system in place for the faculty to evaluate their academic deans; however, the instrument utilized has not had validity or reliability studies conducted on it (D. Scroggins, personal communication, July 14, 2005). Rasmussen (1978) developed an instrument to evaluate academic deans, but again there were no reliability or validity studies conducted for this instrument. Other examples of institutions that have created their own instruments, but without validity or reliability studies, can be found at the University of Michigan, Illinois State University, Texas A & M-Commerce, and the University of Missouri-Kansas City (Frank, 1994; Administrator Evaluation, 2000; Texas A & M University-Commerce Procedures, 1996; UMKC Faculty Senate, 1996).

The position of the academic dean is one that should undergo performance evaluation and for which few valid and reliable instruments exist (Bowker, 1982). Studies that validate an evaluation instrument to assess the performance of academic deans would provide an opportunity to extend the research that has been done on

administrator performance appraisal and to identify factors related to the effectiveness of deans.

In conclusion, performance appraisal of university administrators is an area of interest to both supporters and critics of higher education institutions, especially given the climate of accountability that surrounds colleges and universities. Academic deans are a distinctive group of administrators who frequently must operate within a turbulent environment as they encounter demands from many constituent groups. Further, cultivating support from these constituent groups is critical due to the philosophy of shared governance that exists in higher education. Performance appraisals can assist with determining the strengths and weaknesses of the performance of deans so that they may make improvements and continue to gain the support of their constituents.

STATEMENT OF PROBLEM

The problem that this study seeks to address is that few instruments with demonstrated validity or reliability exist for the important work of evaluating administrators, particularly academic deans, even though formal administrator performance appraisal evolved during the mid to late 20th century. For any institution purporting to measure human performance, it is important that the instrument measure what it is supposed to measure and to do so consistently. Many institutions develop their own evaluation instruments for use in an administrative evaluation system, but the validity and reliability of these instruments have not been established. A review of the *Dissertation Abstracts International (DAI)* database did not yield any performance evaluation instruments that have been developed for the evaluation of academic deans.

Academic deans need to be evaluated so that they may have the opportunity to know and to meet the expectations of their faculty and supervisors (Ory, 2000).

Evaluations also provide a chance to promote communication between the dean and his or her supervisor as well as the faculty, to define or redefine his or her job description, and to determine the perceptive judgment of the faculty toward his or her performance (Wilson & Kroesser, 1981). While academic deans are being evaluated at most institutions, the instruments being utilized do not have established reliability or validity (Bowker, 1982).

This author found only one instrument that has been commercially marketed as a performance evaluation instrument for administrators, and it focuses specifically on the evaluation of department heads. Hoyt developed the Departmental Evaluation of Chairperson Activities for Development (DECAD) system in 1976, which included a form for faculty to rate the importance of the chair's responsibilities as well as the faculty's perceptions of the chair's performance (Knight & Holen, 1985). The DECAD system was updated in 1998 and the name changed to the IDEA Feedback system (A. Gross, personal communication, April 21, 2003). At that time, the IDEA Feedback for Deans and the IDEA Feedback for Administrators were also introduced. However, the psychometric characteristics of the IDEA Feedback for Deans has not yet been established, as the researchers have not collected enough data to make a thorough study (D. Hoyt, personal communication, April 21, 2003; S. Garvin, personal communication, June 24, 2005).

This study will utilize an existing instrument that was originally developed at the University of Tennessee and which has been renamed as the *Administrative Effectiveness*

Survey (AES) instead of the IDEA Feedback for Deans' instrument or any other institutional instrument. The reasons for the selection of the AES are: the IDEA Feedback for Deans' instrument is longer and more complex than the AES and does not probe value disposition of the leader as extensively as the AES.

Birnbaum (1992) explained that good leadership was defined by a leader's effectiveness in making values and activities meaningful to others. Bogue (1992) asserted that values, such as dignity, courage, trust, and responsibility, were key components in the quest for quality. The *Administrative Effectiveness Survey* utilized in this study, therefore, included items related to values of the academic deans and how well he or she upheld those values in his or her actions. The IDEA Feedback for Deans does not include items directly related to values and morals, important dimensions associated with leadership effectiveness, and the instrument is incomplete in its application to evaluate the performance of academic deans based on the definition of effectiveness and quality as proposed by Birnbaum and Bogue.

Other institutional instruments that the researcher found were either too long, which the researcher felt would inhibit respondents from completing the instrument, did not include items regarding the performance of an academic dean, or did not include any items related to the values and morals of academic deans.

In summary, few instruments exist, either institutional or commercial, that have established reliability and validity and that can be used to evaluate the performance of administrators, specifically deans. Academic deans need to be evaluated in order to garner feedback on their performance and while most are being evaluated within some kind of process at their institutions, the instruments utilized are not valid or reliable.

Perhaps one reason for the lack of such instruments can be found in the words of Fox (1969) who indicated that support was needed for more research to develop and validate new data-gathering instruments. He wrote, “Instrument development is such a difficult and time-consuming process that researchers prefer to avoid it whenever possible” (p. 55).

PURPOSE

The purpose of this study was to revise and to establish the reliability and validity of the *Administrative Effectiveness Survey* and to field test the survey on the performance of academic deans at selected institutions of higher education.

RESEARCH QUESTIONS

This study was specifically intended to examine the following questions:

- 1) To what extent are the items in the *Administrative Effectiveness Survey* seen as valid measures of administrator performance effectiveness as determined by two expert panels?
- 2) To what extent does a factor analysis of the item responses in the *Administrative Effectiveness Survey* reflect the initial scales of the instrument?
 - a) Is there any difference in the evaluative factors that emerge across institutional type when classified by Carnegie classification?
 - b) Is there any significant difference when the factor scores are compared to the overall rating score?
- 3) What level of reliability is found for the *Administrative Effectiveness Survey* when the Cronbach Alpha coefficient is calculated?

SIGNIFICANCE OF THE STUDY

This study was designed to establish the reliability and validity of the *Administrative Effectiveness Survey* instrument that may be used in the evaluation of academic deans and that may be adapted for the performance evaluation of other college and university administrators. Most institutions have created their own systems of evaluating academic deans and those instruments do not have established reliability or validity. Valid and reliable instruments are preferable because they provide proof that the instrument is actually measuring what it purports to measure as well as indicating whether the instrument is reliable in multiple test administrations.

This study's findings will suggest a set of criteria to be used in the evaluation of the effectiveness of academic deans. This set of criteria can help academic deans understand what is expected of their performance, rather than operate in a vacuum of unstated expectations. This study's findings may also assist academic deans and their supervisors by specifying areas of performance that need improvement and have implications for training and development. Finally, this study contributes to the existing literature about administrator performance appraisal, in particular providing a reliable and valid instrument that colleges and universities may use to appraise the performance of their academic deans.

ASSUMPTIONS

The basic assumptions for this study were:

- 1) The faculty members who complete the survey will rate their academic dean's performance honestly and will not display bias in their interpretations of the items.

- 2) The faculty members of each college are qualified to provide an evaluation of their academic dean.
- 3) The faculty members are willing to respond accurately and in a timely manner.
- 4) The faculty members have no vested interest in the results of the study.

LIMITATIONS AND DELIMITATIONS

The limitations for this study were:

- 1) The survey instrument will be limited to the criteria identified on the *Administrative Effectiveness Survey* and verified by a panel of experts.
- 2) The sample may not be gender or racially diverse.
- 3) The findings may be subject to interpretations other than those discussed in this study.

Delimitations to the study were:

- 1) The sample of respondents will be limited to faculty members from four different types of institutions as defined by the Carnegie classification system: Doctoral/Research Universities-Extensive; Doctoral/Research Universities-Intensive; Master's Colleges and Universities I; and Master's Colleges and Universities II.
- 2) The sample of respondents will be limited to faculty members from the following colleges: Arts and Humanities, Education, Business, Engineering, Agriculture, Nursing, Law, Social Work.
- 3) Faculty members in this study will be using this instrument to evaluate the performance of their academic deans and not the performance of any other administrator.

Given these delimitations, the results cannot be generalized to other colleges, leaders, and/or situations.

DEFINITIONS OF TERMS

Academic Dean: An academic dean serves as the head of an academic college or school at a public or private university (Morris, 1981). The dean may serve as a member of the president's staff while also holding a faculty position. The dean manages the faculty in his or her college and provides leadership to the college.

Administrator Effectiveness: Administrator effectiveness "...can be defined as the extent to which the administrator performs assigned and expected responsibilities and tasks" and can also be defined by "rating the administrator's style—namely, how does an administrator behave in carrying out his or her responsibilities" (Ory, 2000, p.189).

Evaluation: Evaluation is the utilization of information gathered through measurement techniques for institutional and individual improvement. Astin (1993) described evaluation as having "to do with motivation and the rendering of value judgments" (p. 2).

Formal Performance Appraisal: Formal performance appraisal refers to systematic evaluation procedures.

Performance: Performance "...consists of a performer engaging in behavior in a situation to achieve results" (Mohrman, Resnick-West, Lawler, 1989, p. 49).

Performance Appraisal: Performance appraisal is "(1) ...a formal discussion between a superior and a subordinate (2) for the purpose of discovering how and why the subordinate is presently performing on the job and (3) how the subordinate can perform more effectively in the future (4) so that the subordinate, the superior, and the organization all benefit" (Lefton, Buzzota, Sherberg, & Karraker, 1977, p. 2).

Performance Review: Performance review is “...the activity by which the organization determines the extent to which the employee is performing the job effectively” (Barrett, 1993, p. 159).

CHAPTER II

REVIEW OF THE LITERATURE

This chapter provides an overview of the relevant research and literature that provides the basis for this study. Evaluation of educational policies, programs, and personnel has flowed from both externally and internally mandated accountability pressures. Administrator evaluations constitute one form of evaluation that came from the accountability movement. Accountability in higher education included an accent on policies, programs, and personnel. This chapter first presents a look at the emphasis on accountability policy, then moves to personnel evaluation, including faculty, administrators, and academic deans, and concludes with related literature on existing evaluation instruments.

ACCOUNTABILITY

The period around 1970 is pivotal in the history of higher education accountability policy. Richard Ohmann (2000) described the period between 1945 and 1970 as a time of expansion and prosperity for students and the economy. In particular, one central goal of American higher education has been the goal of enhancing access (Bogue & Aper, 2000). Although higher education remains a privilege, unlike K-12 education, Americans took several steps so that those who had the ability and desire to attend college were able to do so. Comprehensive community college systems were established; and, at state university campuses, programs and services were dramatically expanded. Both state and federal governments made an investment in financial aid for students who wanted to attend college. The Higher Education Act of 1965 provided for federal funds to be designated into financial aid programs. Access was also enhanced

through legislation to stop racial discrimination. The Morrill Act of 1890 prohibited discrimination on the basis of race and led to the establishment of many separate land grant colleges for African American students (Bogue & Aper, 2000). However, this racial segregation was overturned by the 1950s *Brown v. Board of Education* decision, creating access into all colleges and universities for students of all races. Increased access led to an increasingly diverse student clientele attending colleges or universities. After 1970, public funding was less certain. Furthermore, the civil rights, black power, women's liberation, antiwar, and student power movements of the 1960s disrupted some of the institutional postwar complacency at many institutions. As part of a complex reaction to these events, calls for accountability in higher education became more prevalent (Ohmann, 2000).

One of the oldest measures of accountability in higher education is accreditation, which was developed as a method to achieve some measure of quality assurance in higher education. Bogue and Saunders (1992) claimed that accreditation also served "a public need by helping to define institutional mission...and to assist the general public in recognizing and appreciating quality" (p. 36). The North Central Association of Colleges and Secondary schools developed the first set of standards in 1909 to be used in accrediting institutions of higher learning. The most noted types of accreditation are institutional and program accreditation (Berdahl & McConnell, 1999). Berdahl and McConnell (1999) wrote:

Six regional agencies are responsible for accrediting entire institutions' schools, departments, academic programs, and related activities. Program accreditation, extended by professional societies or other groups of specialists or vocational associations, is extended to a specific school, department, or academic program in

such fields as medicine, law, social work, chemistry, engineering, and business administration (p. 81).

The allocation of some federal and state aid to postsecondary education requires colleges and universities to be evaluated by nationally recognized accrediting agencies and provides a measure of accountability. For example, in order for colleges and universities and their students “to participate in Title IV Federal Financial Aid programs, accreditation must be attained from an approved accrediting agency” (Bogue & Hall, 2003, p. 24). Only institutions and programs that are accredited by federally recognized accreditors are eligible for federal funds such as student grants, loans, and research (Harclerod & Eaton, 2005).

Another example of an accountability policy comes from part of the 1992 Reauthorization of the Higher Education Act of 1965, the State Postsecondary Review Entity Program (SPRE). The SPRE program was designed to provide a mechanism for state governments to help the federal government in its oversight of higher education institutions that participated in Title IV programs (Lovell, 2000). Although this program was later terminated by the 1998 Reauthorization of the Higher Education Act, many of the initiatives of the SPRE program can still be found in state-accountability efforts. A report by the U.S. Education Department in March 2002 called for colleges to report their performance on measures to retain students and to graduate them in a “timely fashion” (Burd, 2002, p. A24). Some college leaders and lobbyists believe that this report is reminiscent of the SPRE program and argue that “the quality of higher education...is the responsibility of the institution...and of accrediting agencies, and for public colleges,

state governments” (Burd, 2000, p. A24). William D. Hansen, the Education

Department’s deputy secretary, wrote:

Federal initiatives currently provide over \$70-billion in student aid and direct support of postsecondary education. The department would not be acting responsibly in ensuring the appropriate and proper use of taxpayer funds if it did not consider issues associated with the quality and accountability among all educational institutions receiving federal support (Burd, 2000, p. A23).

Accountability is no longer an internal responsibility of colleges and universities, but rather state, regional, and national stakeholders are creating accountability policies that affect higher education institutions.

In the early 1990s, the role of the state government changed with regard to accountability at colleges and universities (Robles, 1998). Federal funding decreased during the 1970s and 1980s, and state support increased in the 1980s, although the 1990s saw limited state allocations to higher education as mandates for healthcare funding and increased financial support for K-12 education competed in state budgets (Ewell, 1994). State governments began asking questions about the outcomes of higher education, which led to a change in the definition of accountability, from one of access and efficiency to one based on those measures as well as on learning outcomes (Robles, 1998). The policy tool most frequently used to press state concerns about the quality of colleges and universities was student assessment, and by 1990, policy mandates in two-thirds of the states required colleges and universities to assess student learning (Ewell, Finney, & Lenth, 1990).

State policy makers then moved toward an emphasis on the use of performance indicators. Performance indicators emerged in response to pressure for institutions to demonstrate results for resources that were provided by the state (Gaither, Nedwek, &

Neal, 1994). These indicators may serve several purposes: “as evidence of quality, as guides for decision making, as instruments of accountability, as intelligence to track the progress and health trends of an operation, and as signals of goal achievement” (Bogue & Hall, 2003, p. 187). In their 1994 monograph entitled *Measuring Up: The Promises and Pitfalls of Performance Indicators in Higher Education*, Gaither, Nedwek, and Noel provided a detailed discussion of the uses and shortcomings of performance indicators at institutions of higher education both in the United States and at international institutions. In addition, Gaither, Nedwek, and Noel (1994) described a new pattern of performance indicators that emerged in the 1990s and that included: concern with both quality and efficiency, quick implementation with little prior development, deliberate links between planning, budgets, and indicators, comparisons of institutional performance, and policy maker’s resistance to institutional pressure. Performance indicators, though, can be difficult to put into place if the institution’s goals are not known and made clear. Further, it can be difficult to implement the indicators adequately if the indicators are poorly designed. In some cases, the sheer number of indicators that institutions have chosen to employ may confuse decision makers. Gaither, Nedwek, and Noel also mentioned difficulties with the issue of establishing the reliability and validity of performance indicators.

Performance indicators provided the foundation for another rising system of accountability in higher education, that of performance funding (Bogue & Hall, 2003). Accountability moved from an accent on assessment during the 1980s to the linkage of funding and performance in the 1990s (Burke & Seban, 1998). Several states, however, have entered into and moved away from performance funding, such as Texas, Arkansas,

Colorado, and Kentucky (Burke & Seban, 1998; Bogue & Hall, 2003). One exception is the state of Tennessee, which has had a performance funding policy in place since 1979-80 and it is still in effect (Bogue & Hall, 2003). The original Tennessee Performance Funding Policy included an accent on improvement; however, that accent shifted to comparative evaluations based on national or peer norms as well as an acknowledgement of the utility of education decision making (Bogue & Hall, 2003). Performance funding represents a new approach to budgeting that supports a trend toward greater public accountability.

Twenty-seven states agreed in a 1990 Education Commission of the States (ECS) survey that a state assessment initiative was in place as a result of legislative or board action (Callan, 1993). In addition, the American Council on Education reported that 70% of public institutions of higher education said the institutions were assessing learning outcomes, primarily in response to a state requirement. State policy makers are using these accountability demands with the view that higher education needs to be more responsive to societal conditions and state priorities, and more publicly accountable to a broader constituency. In addition to performance indicators, “several states, mostly in the South, publish statistical ‘report cards’ rating the performance of their public campuses” (Peters, 1994, p. 17). These report cards are relatively new phenomena as accountability tools, and so their usefulness has yet to be determined.

Issues of accountability have spurred the evaluation movement in our colleges and universities over the last 30 years. As Bogue and Aper (2000) stated:

We are inclined to see accountability as a formally expressed expectation – a campus or board policy, state or federal law, or formal standard of another agency such as an accrediting agency – that (1) requires evaluation of both administrative

and educational services; (2) asks for public evidence of program and service performance; (3) encourages independent/external review of such performance evidence; and (4) requests information on the relationship between dollars spent and results achieved (p. 214).

No longer are accountability measures focused solely on institutional services and academic programs. Administrative and personnel evaluations have become a part of the institutional responsibility to meet accountability demands of internal and external stakeholders.

PERSONNEL EVALUATION

Faculty

Prior to the 1970s, most of the methods of faculty evaluation were informal. The Purdue Rating Scale of Instruction, developed in 1926, provided an exception as the first published student rating form (Miller, 1972). Extensive research on the Purdue Rating Scale suggested that, “A third of a century of use...by many teachers and a very considerable amount of experimental research...have demonstrated that student evaluation is a useful, convenient, reliable, and valid means of self-supervision and self-improvement for the teacher” (Remmers & Weisbrodt, 1965, p. 1). Teaching evaluations were initiated on many college campuses in the 1960s as students requested a voice in their education. In the 1970s, economic changes, faculty demands for a voice in institutional governance, and the emphasis on accountability created the formalization of faculty evaluation.

Methods of faculty evaluation include student rating forms, peer review, self-evaluation, and administrative evaluation; however, much of the recent research has been conducted specifically on student rating forms (Whitman & Weiss, 1982). While many

scholars agreed that these forms were a good measure of a professor's skills, Cahn (1994) disagreed and claimed that the rationale for the use of student evaluations was flawed if based upon the belief that "...when faculty members fail to fulfill their obligations, students suffer the consequences" (p. 37). He explained, for example, "When airplane pilots fail to fulfill their obligations, passengers often suffer the consequences, but it does not follow that passengers should have a strong voice in evaluating pilots" (p. 37). Nevertheless, nearly 2000 studies have been completed on the topic of student evaluations of faculty performance (Wilson, 1998). As of 1973, about 30% of colleges and universities used student evaluation of faculty members, but almost 100% use them today. Wilson (1998) said, "Student ratings carry more and more weight, especially on campuses where the focus is on teaching. Such evaluations are now the most important, and sometimes the sole, measure of an instructor's teaching ability" (p. A12). McKeachie and others (1971) concluded from their study of the literature on the validity of student ratings that student ratings did have some validity, and students who rated teachers as effective tended to be those students who learned the most.

Although the history and current status of faculty evaluations appear to be heavily dependent upon student ratings of courses, there is some trend toward the expanding use of peer review and post-tenure review as components to faculty evaluations. Teaching has not necessarily been seen as a scholarly activity, and as such, resistance to peer review has traditionally been strong (Licata, 1999). Boyer (1990) wrote that, "teaching is often viewed as a routine function, tacked on, something almost anyone can do" (p. 23). Boyer argued that, while the dominant view of a scholar was to be a researcher, scholarship needed to be redefined in order to more accurately describe the work of the

professoriate as “the scholarship of *discovery*; the scholarship of *integration*; the scholarship of *application*; and the scholarship of *teaching*” (p. 16). Bogue and Aper (2000) posited that the use of teaching portfolios may help bring the art of teaching into the public forum. In response to claims that summative faculty evaluation does not lead to improved teaching, some scholars have suggested that formative peer evaluation be developed and implemented (Keig & Waggoner, 1994). Methods that have been used in formative peer review include: “direct classroom observation, videotaping of classes, evaluation of course materials, as assessment of instructor evaluation of the academic work of students, and analysis of teaching portfolios” (Keig & Waggoner, 1994, p. v).

Many institutions have adopted the process of post-tenure review, which usually includes significant peer review and career planning. Licata (1999) stated that 37 states had adopted system-wide post-tenure review policies, implemented them in selected colleges and universities, or were considering the process.

Administrators

Early accountability measures involved the evaluation of students and faculty and have expanded to include administrator evaluation. Wattenbarger (1983) stated, “It appears illogical...for a college to develop procedures to evaluate programs, to measure student progress, to evaluate faculty and career service staff, while ignoring the need to establish procedures for evaluating the executive and management staffs” (p. 46). After examining practices in administrator performance appraisal systems, Lahti (1978) described the common systems as: 1) global essays and ratings written by the supervisor; 2) trait or personality rating scales; 3) peer rankings; 4) organizational records such as accident or absence rates or sales figures; 5) critical incidents which utilize descriptions

of actual behavior as items for measurement; 6) behaviorally anchored rating scales (BARS) which consist of the creation of measures that are job-relevant and not goal-oriented; and 7) management-by-objectives (MBO) which consists of the manager, with or without the employee, setting objectives for a set period and then evaluating progress at the end of that period. The most often used method in organizational settings is the unstructured narration or essay method (Lahti, 1978).

Some administrative evaluation systems utilized a bottom-up approach where faculty members were asked to evaluate those administrators above them in their academic unit. Morris (1981) claimed that this type of system invited criticism by those who believed that faculty members knew very little of the demands of the dean. According to Morris (1981), "Criticism of administrators as a class is common practice. Hence, the evaluation of superiors only formalizes what most faculties would consider their routine, institutional duty" (p. 147). Belcher (1992) wrote about a bottom-up evaluation system in place at a community college. A committee defined administrator excellence and used those statements to create a 24-item survey. Faculty rated administrators on the 24 items. As a group, the administrators had definite strengths and weaknesses. The items with the highest mean ratings were: knowledgeability of policies and procedures, recognizing that the first constituency to be served is the student, and knowledgeability of the characteristics of the student body. The items with the lowest mean ratings were: deals effectively with inappropriate behavior in a timely manner and emphasizes importance of student advisement. A factor analysis on the data from the 24 items revealed that the respondents viewed the administrators across five dimensions: general administrator competence, interpersonal skills, support of college mission,

knowledge base, and motivator. Although Belcher wrote that the system was piloted twice, nothing was said about the reliability of the survey. This instrument also did not have any items related to the values of administrators.

Wilson and Kroesser (1981) described the development of a similar administrative evaluation system at the West Virginia College of Graduate Studies. A committee comprised of five administrators and two senior faculty members developed a comprehensive administrative evaluation system that included a narrative self-evaluation for each administrator, an administrative office evaluation, and an evaluation of individual administrators. After reviewing literature and evaluation instruments used at other institutions, the committee created an instrument with eight general criteria that was used to evaluate individual administrators. The eight criteria included items such as: communicates effectively with others, is accessible and helpful, demonstrates good judgment and consistency in decision making, and exhibits fairness in dealing with others. The statements on the instrument primarily dealt with interpersonal relations, planning functions, and efficiency and used a 5.0 scale (poor to excellent). Faculty, staff, and administrators were sent a packet of evaluation forms to fill out for each included administrator's name. The packet contained a randomly selected group of administrators to be evaluated. Wilson and Kroessner (1981) reported that, "The overall or composite mean for all administrators combined was 3.56 (average to above average). The individual administrator means ranged from the mid-two area (below average) to over four (above average to excellent)" (p. 47, 49). There was no discussion in this article about the validity or reliability of the instrument. This instrument also did not contain any statements pertaining to the values of the administrators.

One method of assessing the effectiveness of administrator performance is to determine whether the administrator enjoys the support of the institution's strategic constituencies (Birnbaum, 1992; Fujita, 1990). Birnbaum (1992) based *How Academic Leadership Works* on data collected during a five-year longitudinal study of how college and university presidents and other leaders interact and communicate. As a part of that study, Fujita (1990) interviewed campus leaders at 32 campuses to determine what they had to say about the quality of presidential leadership on their campuses. Content analysis revealed 10 evaluative dimensions:

1) willingness to be influenced by others, 2) competence, 3) respect for the culture of the college, 4) encouraging the free flow of information, 5) association with outcomes, 6) involvement, 7) commitment to the institution, 8) a leader primarily through position, 9) appearance, image or impact, and finally, 10) a comparison with the predecessor (p. 5).

Trustees and administrators usually rated their presidents favorably, whereas faculty members were more divided in their responses. The role of the evaluator (trustee, administrator, or faculty leader) was a significant variable. Data revealed that faculty ratings and full constituency support, which included the three constituency groups (trustees, faculty, and administrators), were highly correlated. The most significant measure of constituent support was deemed to be faculty backing. Based on results from Fujita's study (1990), Birnbaum (1992) determined that college presidents must be simultaneously responsive to three major constituency groups: trustees, faculty, and administrators.

Administrator systems of evaluation seem to be fairly prevalent in institutions of higher education (Lynch & Bowke, 1984). In the 1980s, the most common method of administrative evaluation consisted of a casual implementation of the management by

objectives process (Stronge & Helm, 1991). However, most colleges and universities now utilize rating scales to evaluate administrators (Miller & Miller, 2000; Belcher, 1992). Miller and Miller (2000) asserted that most evaluation systems are rarely evaluated.

Academic Deans

An academic dean is one university administrator whose position has not been studied as much as other university administrators, such as the president or department chair (Dill, 1980; Bowker, 1982; Martin, 1993). This is particularly distressing given the importance of the role that the dean holds at most institutions of higher education. Historically, the position of the dean goes back to the 1860s when most colleges or universities only had four administrative officers (Dill, 1980). A dean for the Harvard medical school was named in 1864. By 1900, fewer than 20% of colleges and universities had named deans, but deans became more prominent by the 1940s and were sometimes viewed as extensions of the president. Dill (1980) stated, “By the 1960s, deans had ‘arrived’. There were more on some campuses than there had been faculty members a century earlier” (p. 624). For example, in a 1976 study, 1138 institutions, of which 705 were four-year colleges, were surveyed to determine how many deans or directors of colleges existed (Dill, 1980). The study found that 416 colleges had deans or directors of colleges of arts and sciences, 392 had deans or directors of colleges of business, and 338 had deans or directors of education. In large universities, deans were seen as a primary force for change and innovation. Even with these increasing numbers and influence, researchers have not extensively studied deans.

Much of what has been written about academic deans has been descriptive in nature, dealing largely with their roles and responsibilities, traits and backgrounds, and governance and decision-making styles (Martin, 1993). For example, Enarson (1962), an early researcher, delineated the roles of an academic dean as: budgeteer, personnel officer, academic planner, innovator, coordinator, lightning rod, and majority leader. More recent authors have described the roles of a dean as leader, manager, scholar, faculty developer, and administrator (Sarros, Gmelch, & Tanewski, 1998; Wolverton, Montez, & Gmelch, 2000). Other authors have written about skills and experiences that academic deans should have in order to be successful; unfortunately, almost none of their work has been tested by research or developed from a research base (Miller & Miller, 2000). Such skills include experience in administration, supervision, budgeting, program assessment, faculty development, ability to bargain, integrity, fairness, confidence, and competence (Robillard, 2000; Scott, 1979).

Some authors claim that the success of colleges and universities will depend on the dean's performance and ability to meet the challenges facing colleges and universities in the future (Bok, 1986; Wolverton, Montez, & Gmelch, 2000). One study utilized the Dean's Task Inventory (DTI) to assess the importance of 32 tasks of the dean's role (Wolverton, Montez, & Gmelch, 2000). The study surveyed deans at 360 institutions to determine what they saw as important typical tasks of a dean. A factor analysis was performed on the data, resulting in six dimensions that represented the primary roles of a dean: external and political relations, leadership, resource management, internal productivity, personal scholarship, and academic personnel management. The DTI was based on the Chair Tasks Inventory which has had its reliability and validity established

through subsequent research (Sarros, Gmelch, & Tanewski, 1998). The reported Cronbach alpha for the Dean's Task Inventory is 0.89. The DTI was not developed as a performance evaluation instrument and has not been utilized as an evaluative instrument. Further, the items on the instrument dealt with tasks and not values of the academic dean. For those reasons, the DTI was not selected over the AES for this study.

Rasmussen (1978) developed an instrument for assessing the performance of deans. He proposed an evaluation instrument with 22 items that would be used by the dean, the faculty, and the president. The items were described in behavioral traits such as: administrative management, executive judgment, delegating authority and responsibility, providing academic leadership, acting decisively, planning ability, encouragement of faculty research, and improvement of teaching. Each group of evaluators would assess the dean's performance, responding with a rating of the ideal performance of the dean as well as the actual performance of the dean. Rasmussen did not include any discussion of where the items came from or the validity or reliability of the instrument, nor were there value statements included in the list of traits for evaluation.

Two studies reported attempts to determine what behaviors of deans constituted effective performance. Earlier research focused on what deans should do to be effective (Dill, 1980; Sarros, Gmelch & Tanewski, 1998); whereas Martin (1993) and Skipper (1982) reported what effective deans actually do and how they are perceived to be successful. Martin (1993) asked five vice presidents, provosts, chancellors, or presidents to nominate an effective dean at their respective institutions. Using a multiple case study design, Martin interviewed those deans and other individuals who could talk about those deans. Subsequently, Martin identified themes related to college community members'

understanding of effective dean's leadership. Martin concluded that effective deans have five areas of expertise: cultural representative, communicator, manager, planner/analyst, and advocate. Effective deans represented the culture of their colleges and universities by believing in and working for "the cause" of the institutions. Effective deans also had a strong commitment to and continually worked on communication processes as well as being skilled managers by offering ideas and knowing how to implement those ideas. Martin stated that effective deans engaged in collective vision-building as well as building linkages outside the college in order to strengthen their advocacy roles. Martin's study was qualitative in nature and did not include a survey instrument that could be compared to the AES. However, the themes that Martin found in her research are represented in the AES.

In an attempt to develop a tool to potentially help search-screening committees as they interview candidates, Skipper (1982) conducted a study to determine whether four indicators of administrative effectiveness could discriminate between effective and ineffective deans. Previous research by Skipper (1976, 1977) identified 10 personal characteristics and seven administrative skills, which were subsequently reduced to the four indicators used in Skipper's 1982 study. The four indicators were: intellectual efficiency, flexibility, knowledge of position, and judgment. Eighteen university presidents and provosts were asked to rate a "poor" academic dean as compared to other academic deans on their campuses, and an additional 18 presidents and provosts were asked to rate an "outstanding" academic dean. By using scores of discriminate functions for each variable, an equation was established to predict a leader's location on a continuum for administrative effectiveness. The equation was Administrative

Effectiveness = (.20 Intellectual Efficiency) – (.14 Flexibility) + (.16 Knowledge of Position) + (.14 Judgment) – 2.07 and was referred to as the administrative effectiveness index. The results of Skipper's study suggested that the administrative effectiveness index was accurate in identifying ineffective deans, but not as accurate as the index predicted for identifying outstanding deans. Skipper concluded that if search-screening committees focused on the four indicators and candidates scored low on them, the committee could then rule out the candidates for future consideration. Skipper's study provided another example of research that determined what effective deans do. However, his results were more indicative of what ineffective deans do. The items in Skipper's study have not been utilized as an evaluative instrument, but Skipper's four indicators are similar to some of the items on the AES.

Several articles have been written about proposed criteria for evaluating academic deans. Morris (1981) described two dimensions of evaluation criteria as vital to the evaluation of deans and which should be given weight in order to improve the dean's administration of his or her college. One dimension, style, meant measuring the dean's personal manner of going about the daily tasks of leadership. Style included such items as "...accessibility, warmth, openness, emotional control, fairness, listening ability, decisiveness, firmness under pressure" (Morris, 1981, p. 149). Style also encompassed how much the dean consulted with others and the dean's perception of the job, whether the dean would be an ultimate, solitary decision maker or a broker and arbitrator. The second dimension, results, meant determining the quality of the institution that the dean's activities brought forth. Results come from a variety of sources, such as:

...the national reputation of the faculty, the caliber of teaching they are known by, the amount of external funding they generate, what the accrediting agencies say, the overall estimate of outsiders from comparable or better institutions, the employability of its graduates, the testimony of those who benefit from its community services or who employ its alumni (Morris, 1981, p. 152-153).

Morris (1981) recommended criteria that an evaluator might use as a basis for evaluating a dean: “1) ...leadership of the unit, 2) leadership to related agencies in the external world, 3) a talent finder and builder of the unit, 4) planning ability, 5) accessibility and communication with various constituencies, 6) resource allocation, 7) overall effectiveness, i.e., output of the unit” (p. 169). Although Morris did not provide any examples of anyone using these criteria, many of these criteria are represented as items on the *Administrative Effectiveness Survey* utilized in this study. Morris also did not have any items related to the values of the academic dean which are included on the AES.

After reviewing relevant literature, Harada (1991) proposed four classes of criteria for faculty to use these criteria in their evaluation of the dean for improved administrative performance. The four classes of criteria were: knowledge, managerial skills, leadership skills, and attitudes and personal attributes. Knowledge referred to specific information that the dean should have in order to adequately perform his or her job duties. Management defined the functions that promote and continue the bureaucracy of administration; whereas, leadership defined those skills needed to create goals and vision. Attitudes and personal attributes referred to worldview, values, and character of the dean. Harada (1991) did not suggest that faculty members evaluate the dean with respect to all items within each class of criteria, claiming that the faculty did not have adequate knowledge to judge a dean’s performance on all aspects of the dean’s job. However, Harada’s study provides support for the use of faculty evaluating deans and

many of the items on the AES are reflected in the criteria found in Harada's review of the literature, particularly those items involving values of the dean. Harada's study did not include evaluative items and did not produce an evaluative instrument.

Lasley and Haberman (1987) surveyed Vice Presidents of Academic Affairs to determine what criteria they used to evaluate deans of colleges of education and how important each criterion was. A questionnaire comprised of 20 items was mailed to each Vice President for Academic Affairs of the National Association of State Universities and Land Grant Colleges. The 20 items were divided into 6 sections: decision making, communication, faculty, professional values, change, and affirmative action. The two items assessed as most important for evaluative criteria for deans were: "...the dean chooses and supports quality faculty" and "...the dean maintains quality standards in his/her recommendations for promotion" (Lasley & Haberman, 1987, p. 14).

Maczynski, Lasley, and Haberman (1989) extended this research by surveying faculty members to identify criteria that the faculty members believed to be important in the evaluation of their deans. The authors mailed a different questionnaire to 500 members of the American Educational Research Association and 231 were returned. The authors chose this population of research-oriented faculty because of the increased emphasis on research and publications and to ascertain how those faculty members who are engaged or interested in research evaluate their deans. The questionnaire included 49 questions divided into 8 subsections: supervisory skills, knowledge of teacher education, problem solving skills, communication skills, personnel management skills, resource management skills, affirmative action skills, and academic standards skills. The respondents ranked communication skills as the most important factor and affirmative

action skills as the least. Other important items included bargaining effectively with superiors for resources, maintaining quality standards in recommendation of individuals for tenure and promotion, recruiting quality faculty, nurturing academic growth, and cooperating with faculty to find solutions to identified problems. This instrument was developed to identify the factors that faculty used to evaluate their academic deans and so respondents rated the items with regard to the item's importance in evaluation of academic deans. The respondents were not actually evaluating the performance of their academic deans. The authors of the study field tested the instrument with a group of experts to determine internal validity; however, no mention was made of any reliability statistics. This instrument was not selected for use in this study because of the length of the instrument and the fact that none of the items probed values of academic deans.

The studies discussed here have addressed the roles of the dean (Dill, 1980; Enarson, 1962; Sarros, Gmelch, & Tanewski, 1998), evaluation systems (Rasmussen, 1978), and types of criteria that should be used when evaluating deans (Martin, 1993; Morris, 1981; Harada, 1991; Matczynski, Lasley & Haberman, 1989), yet few studies discuss how deans are actually rated on their job performance. It seems prudent that colleges and universities would want to know the areas in which academic deans perform well and what areas they may need to develop so that the performance of the institution as a whole may be improved. Although it is difficult to show empirically that the performance of the dean affects the institution, assessing the performance of academic deans should provide a starting point for future study. In order to assess the performance of academic deans, though, valid and reliable instruments are needed. Wattenbarger

(1983) said, “There is a continued need to develop better data collecting instrumentation” (p. 53).

EVALUATION INSTRUMENTS

This portion of the literature review includes studies that describe the development or revision of evaluation instruments. Several instruments that have been developed specifically as evaluation instruments for college and university administrators do not include any information regarding the reliability or validity of the instruments (Rasmussen, 1978; Belcher, 1992; Wilson & Kroesser, 1981). Brown and Rodriguez (1989) provided an exception with the development of a self-evaluation instrument to measure administrative job performance at three community colleges. The purpose of this administrative evaluation system was to assist individual administrators in planning their own self-improvement programs. The instrument was constructed with standards written in the form of behaviorally anchored rating scales (BARS) and contained 30 performance dimensions that had been reported in the literature. The performance dimensions included such areas as: academic planning, accessibility, communication, initiative, leadership, and supervision. Administrators at the three community colleges reviewed the instrument to establish its validity. For each performance dimension to be included in the instrument and to establish content validity, a minimum of 70% of the administrators surveyed had to agree that the dimensions included in the instrument were descriptive of an area of administrative performance. This benchmark was exceeded for each performance dimension. In addition, the administrators unanimously endorsed the rating scales. Reliability of the instrument was established using a test/retest measure. The reported Pearson Product Moment Correlation Coefficient was greater than .50 for

all but nine of the dimensions. The development of this instrument followed suggested survey development practices, yet was designed as a self-evaluation instrument for administrators. This instrument was not chosen for use in this study because it did not include performance dimensions specific to the academic dean nor did it include items related to the values of administrators.

The Departmental Evaluation of Chairperson Activities for Development (DECAD) system, mentioned earlier, is another instrument that was developed and had its reliability and validity established. The DECAD system included two forms, chairperson information (CI), which solicited Likert-type ratings of the importance of fifteen responsibilities, and the second form, faculty reactions to chairpersons' activities (FRCA), which gave faculty's ratings of each responsibility's importance as well as their perception of the chairperson's performance (Knight & Holen, 1985). Hoyt and Spangler (1978) developed the FRCA by revising the Leadership Behavior Description Questionnaire (LBDQ), which had been designed by Halpin (1957). Ten of the 40 items on the LBDQ were excluded because Hoyt and Spangler determined that they were unrelated to faculty ratings of department head's effectiveness. Data were collected from four institutions and 103 department heads were included, with the number of faculty raters for each department head varying from 5 to 30 and averaging 13. "Split-half reliabilities for performance ratings of department heads on the DECAD ranging from 0.60...to 0.91...Intra-class correlations from 0.51...to 0.81...are also reported" (Knight & Holen, 1985, pp. 679, 682). The commercially available instrument, the IDEA Feedback for Department Chairs, is based on the DECAD instrument. Faculty members rate the effectiveness of the department chairs in performing five types of administrative

responsibilities as well as provide an overall judgment of the chair's performance. The current form has some new items that have not yet been empirically tested. The IDEA Feedback for Dean's instrument has not had a thorough study of its psychometric characteristics yet and is the only instrument commercially available at this time for the evaluation of academic deans.

A review of the *Directory of Unpublished Experimental Mental Measures*, volumes seven and eight, did not yield any relevant performance evaluation instruments (Goldman, Mitchell, & Egelson, 1997; Goldman & Mitchell, 2003). A review of *Tests in Print V* produced a couple of related instruments that one might utilize for the evaluation of administrators, but were not designed for higher education (Murphy, Impara, & Plake, 1999). The Employee Effectiveness Profile was designed to assist managers in identifying the overall effectiveness of individual employees (Aleamoni, 1995). Managers rate their employees on a 20-item questionnaire. However, there is no available information on the development, reliability, or validity of the questionnaire.

The Leadership Behavior Description Questionnaire (LBDQ) was designed to elicit employee ratings of a supervisor to describe his or her leadership behavior (Dipboye, 1978). The LBDQ was developed by constructing 130 items which were tested on a questionnaire in a study of bomber crews in the U.S. Air Force. Factor analysis revealed four factors, with initiating structure and consideration accounting for 83% of the variance. Fifteen items loaded on the initiating structure factor, 15 items loaded on the consideration factor, and 10 items were used as buffer items for the final version of the LBDQ. Dipboye stated that the two factors have reported high coefficients of internal consistency as well as high interrater agreement. Halpin (1957) wrote that the

estimated reliability by the split-half method was .83 for the initiating structure factor and .92 for the consideration factor, when corrected for attenuation. Validity was established by finding significant correlations between job satisfaction and work group performance with the LBDQ scales. The LBDQ, though, was not designed to evaluate the performance of academic deans and has not been used as an evaluative instrument. In addition, the items on the LBDQ do not include items pertaining to the tasks of academic deans nor are there any items regarding the values of academic deans which is why this instrument was not chosen over the AES.

A review of the *Dissertation Abstracts International (DAI)* database did not reveal any dissertations that developed an instrument to evaluate the performance of deans. There were a few dissertations with related topic areas. Yang (1987) developed an appraisal instrument to evaluate department chairpersons in Taiwan private junior colleges. A questionnaire was used to determine consensus among junior college personnel groups regarding the department chairperson's roles and responsibilities. The Pearson correlation method was used and a high positive correlation was found between groups, which validated the content of the appraisal instrument. The Pearson correlation method was also used to determine the relationship among the appraisal scores given by superiors, peers, and subordinates of the chairs using the new appraisal instrument. Positive correlations were reported. This instrument did not include evaluative items for the performance of academic deans, nor did it include items related to the value disposition of the academic dean.

Pinero (1988) developed an instrument to measure expectations of the role of the academic dean in Puerto Rico in order to determine whether faculty expectations were

similar or not to the expectations of the deans themselves. Pinero used a review panel to examine the 34-item instrument for content validity. Academic deans from 27 colleges and 120 faculty members were asked to rank the items according to the importance they believe each function should have for the academic dean. The reliability of the instrument was established through a Cronbach alpha test. While this instrument has had its reliability and validity established, it did not include items relating to the values of an academic dean and which were included in the AES.

Montez (2002) developed the Higher Education Leadership Instrument (HELI) to assess five dimensions of leadership: integral, relational, credibility, competence, and direction/guidance. The HELI instrument contained 32 indicators for leadership behaviors and 32 core attributes. The HELI was administered to a sample of 452 deans with a 51.3% response rate. The reported alpha for internal consistency of the instrument ranged from .74 to .91. Principal components analyses tested whether the indicators loaded on their respective dimensions; however, a general leadership factor emerged that accounted for the majority of the test's variance. The HELI was not administered as an evaluative instrument, but rather allowed academic deans to determine what leadership behaviors they utilized. Although this instrument did include items under the dimension of credibility, which are values-based behaviors, this instrument was not selected for use in this study over the AES because of the length of the instrument.

This review of evaluation instruments reveals a void in the literature with regard to the validation of performance evaluation tools used in the evaluation of administrators, in particular of academic deans, at colleges and universities and further supports the purpose of this study.

CONCLUSION

Accountability demands from internal and external stakeholders of colleges and universities have led to more systematic evaluations of administrators. Academic deans are a group of administrators that wield a great deal of influence on college and university campuses. Their skills, duties, and tasks have been identified in the research; however, little has been written to validate criteria to assess their effectiveness. Although some critics claim that faculty members do not know enough to be able to evaluate the performance of academic deans, others refute that claim (Morris, 1981; Birnbaum, 1992). Fujita (1990) determined that the role of the evaluator is a significant variable in distinguishing between good and bad presidents. It seems reasonable that faculty members can evaluate the performance of academic deans. Although there is no general agreement as to which sources of evaluation are more appropriate (Nordvall, 1979), several examples exist where colleges and universities have utilized faculty to evaluate academic deans and other administrators (Budig, 1995; Belcher, 1992; Wilson & Kroessner, 1981). This study proposes that faculty members are a valued source of evaluation for the performance of academic deans.

Very few studies included in this literature review described the development of an evaluation instrument for university administrators, and most of them did not establish the validity or reliability of the instrument (Belcher, 1992; Wilson & Kroessner, 1981; Rasmussen, 1978). Only one commercially available instrument exists for the evaluation of academic deans, and the validity and reliability of that instrument have not yet been determined. This study should contribute to the literature about administrator performance appraisal by providing a valid and reliable evaluation instrument.

CHAPTER III

RESEARCH DESIGN AND METHOD

INTRODUCTION

The purpose of this study was to revise and to establish the validity and reliability of the *Administrative Effectiveness Survey* and to field test the survey on the performance of academic deans of selected institutions of higher education. The research questions for this study were:

- 1) To what extent are the items in the *Administrative Effectiveness Survey* seen as valid measures of administrator performance effectiveness as seen by two expert panels?
- 2) To what extent does a factor analysis of the item responses in the *Administrative Effectiveness Survey* reflect the initial scales of the instrument?
 - a) Is there any difference in the evaluative factors that emerge across institution type when classified by Carnegie classification?
 - b) Is there any significant difference when the factor scores are compared to the overall rating score?
- 3) What level of reliability is found for the *Administrative Effectiveness Survey* when the Cronbach Alpha coefficient is calculated?

An initial small-scale pilot test of the survey was conducted, and then the survey was field-tested nationally at selected institutions as described in this section. The following sections detail information on the instrument revision, the methodology for the initial pilot test, and the methodology for acquiring the data in a national field study to establish the validity and reliability of the instrument.

INSTRUMENT REVISION

This study was designed to be a reliability and validation study of a revised, existing evaluation instrument. The instrument that was used in this study came in large part from an instrument initially developed by Dr. Grady Bogue at the University of Tennessee for the Faculty Senate and entitled, *Faculty Senate Administrative Questionnaire*. Dr. Bogue reviewed the literature on leadership effectiveness and incorporated his own extensive experience as a collegiate administrator to develop the items included on the *Faculty Senate Administrative Questionnaire* (Appendix A).

Several aspects of the *Faculty Senate Administrative Questionnaire* were revised prior to establishing the *Administrative Effectiveness Survey* used in this study. First, within the demographic portion of the instrument, an option for administrative positions was deleted and an option for selecting one of eight colleges was added - Agriculture, Arts and Humanities, Business, Education, Engineering, Law, Nursing, and Social Work. Next, the Likert scale was changed from “strongly disagree, disagree, agree, strongly agree” to either “poor, fair, good, excellent” or “unsatisfactory, needs improvement, satisfactory, exemplary”, one of which would be selected depending on the results of two review panels.

Ory (2000) recommended that evaluations for key administrative positions should have questions crafted to the factors most related to the specific position. Some of Ory’s suggested factors for the academic dean included: providing academic leadership, creating an effective working relationship with the faculty, and maintaining academic credibility. Belcher (1992) had also recommended that an evaluation instrument include items specific to an administrator’s position. After a review of the relevant literature

(Maczynski, Lasley, & Haberman, 1989; Lasley & Haberman, 1987; Sarros, Gmelch, & Tanewski, 1998), a section was added with seven items regarding the performance of academic deans. These items involved tasks of the deans that faculty, deans, and administrators scored as important functions of the dean. For example, Sarros, Gmelch, and Tanewski (1998) asked deans to rank the important duties of a dean and the item “represent the faculty to the administration” was the highest with a mean of 4.77 out of 5.0, thus it was selected for inclusion on the AES. Finally, the two open-ended questions asking for illustrations of exemplary performance or performance needing improvement in the original instrument (see Appendix A) were deleted. A section was added for survey respondents to answer two questions as to whether they would include the open-ended questions and two areas for respondents to provide feedback on the instrument itself. Two versions of the instrument can be found in Appendix B, each with a different response scale.

The initial *Administrative Effectiveness Survey* was divided into three parts. The first part consisted of demographic data, including options to select gender, race, and college. The second part consisted of 32 evaluation criteria for the performance of the dean; for example, “establishes appropriate goals for self”. Twenty-five questions were organized into four areas: leadership planning and goal setting, leadership style and skill, leadership resource stewardship, and leadership values and standards. A fifth area was included with seven questions specific to the position of the dean. A 4-point Likert scale was used, ranging from poor to excellent or unsatisfactory to exemplary, with an option at the end of the scale for ‘no basis for evaluation’. The third part included a question for the overall assessment of performance with a scale of unsatisfactory, needs improvement,

satisfactory, and exemplary. The last portion of the survey contained questions for respondents to make suggestions about the survey instrument.

The revised instrument used in this study, the *Administrative Effectiveness Survey*, followed the criteria suggested by Bourque and Fielder (1995):

- Do not make unrealistic time estimates.
- Ask enough questions to obtain the information needed.
- Use space between questions.
- Use vertical format, space, boxes, arrows, shading, or other devices consistently to maximize the clarity and order of questions.
- Do not avoid precoded response categories, but clearly indicate the code that corresponds to each response.
- Consider the use of simple grids.
- Use a booklet format when possible.
- Have good contrast between print and paper.
- Use 10-point pitch.
- Use an easily read, equally spaced font, such as Courier.
- Avoid italics.
- Use bold, underlining, or CAPITALS judiciously and consistently for emphasis and instructions.
- Do not split instructions, questions, and associated responses between pages (p. 102-103).
- After the questionnaire was constructed and formatted, the next step was to

establish the validity and reliability of the instrument.

Instrument Validity

Validity is an important measure of a survey instrument's utility. Validity is a test of how well the instrument measures what it purports to measure. Litwin (1995) identified four types of validity associated with measuring the performance of a questionnaire: face, criterion, construct, and content.

Face validity is often confused with content validity. However, face validity is the least scientific measure of all the validity measures because face validity is based on

a casual review of the survey's items by untrained judges. Many researchers do not even consider face validity as a measure of validity.

Criterion validity is a measure of how well the questionnaire correlates to another outside instrument or predictor. The survey instrument in question is judged against another method that is acknowledged as established and held to a high standard of being able to accurately measure the same variable. A correlation coefficient is calculated between the two instruments and a high correlation suggests good criterion validity.

Construct validity is a measure of how meaningful the survey instrument is when in practical use. In order to determine construct validity, the instrument is tested for convergence across different measures of the same trait and is tested for divergence between measures of related but conceptually different traits. The instrument should correlate highly with measures that the construct indicates they should as well as correlate less with the measures that the construct indicates they should not.

Litwin (1995) described content validity as, "... a subjective measure of how appropriate the items seem to a set of reviewers who have some knowledge of the subject matter" (p. 35). Content validity is not quantified with statistics; but rather, it is reflected as an overall opinion by a group of trained reviewers. One way to assess content validity is to conduct focus groups with representatives of the target population and collect their comments on the survey's design, content, and relevance. Another way to establish content validity is to ask expert representatives from the field of study to review the instrument and provide feedback.

To ensure the content validity of the questionnaire used in this study, two expert panels were asked to review the instrument and make comments (Appendix C).

However, prior to these two panels, the initial instrument had been reviewed by Dr. Bogue and a panel of six faculty members for content validity and they concluded that the items on the original instrument should remain. The two panels selected in this portion of the study reviewed the revised instrument. The first panel of experts consisted of the University of Tennessee faculty serving on the committee guiding this study: Dr. Grady Bogue, Dr. Norma Mertz, Dr. Bill Lyons, and Dr. Mary Sue Younger. All four of these individuals responded and provided their comments and suggestions. The second panel contained national experts in their respective fields. Members for this panel included: Robert Birnbaum, Professor of Educational Policy and Leadership at the University of Maryland, College Park; Donald Kennedy, President, Emeritus at Stanford University; Marvin Peterson, Professor of Education at the University of Michigan; E. K. Fretwell, Chancellor Emeritus at the University of North Carolina, Charlotte; C. Peter Magrath, President of the National Association of State Universities and Land-Grant Colleges; James L. Fisher, former President of the Council for Advancement and Support of Education (CASE); Kim S. Cameron, University of Michigan, Director of Organizational Studies Program, National Center for Higher Education Management Systems; Richard Miller, Professor of Higher Education at Ohio University; Van Cleve Morris, University of Illinois, Chicago. Five of these individuals responded with comments regarding the instrument, which was a much lower number than the researcher had hoped. Revisions were made to the initial instrument based on the suggestions by those experts.

As a further test of content validity, factor analysis was performed on the data from the national field study. Principal components analysis can first determine the

initial number of factors required for the factor analysis. The factor analysis can then determine the extent to which the instrument items measure common or different content areas. Litwin (1995) explained that, “A factor is a hypothesized trait that is thought to be measured with items in a scale” (p. 47). Factor analysis can determine whether the items exhibit patterns of relationships with each other, such that the items in each factor correlate highly with each other (Johnson, 1998). After validity has been established, the researcher must next establish reliability of the instrument.

Instrument Reliability

Reliability is a measure of how reproducible the survey instrument’s data are across participants in multiple test administrations. Litwin (1995) described two common types of reliability used in survey design: test-retest and internal consistency.

Test-retest reliability is measured by having the same set of respondents complete a survey at two different points in time to see how stable the responses are. A correlation coefficient is then calculated between the two sets of responses. A correlation coefficient of 0.70 or higher is considered good and indicates that the survey is reliable.

Another common measure in assessing survey instruments is internal consistency reliability. Internal consistency is used when the survey is only administered once and serves as an indicator of how well different items measure the same issue. A coefficient alpha is found by finding the variance of all responses for each item and then adding those variances across all items in the scale. If a questionnaire answer scale uses multiple answers: strongly disagree, disagree, agree, and strongly agree, then the Cronbach Alpha coefficient is calculated. If a questionnaire answer scale has only two

answers: yes or no, then the Kuder-Richardson coefficient is calculated. This study employed the calculation of the Cronbach Alpha coefficient in order to measure the reliability of the instrument. The higher the correlation coefficient is, the more reliable the questionnaire should be.

Instrument Pilot Study

A pilot study was conducted in order to allow the researcher to make any necessary changes to the instrument or the data collection method prior to the national field study of the instrument.

Prior to the pilot study, two panels were asked to review and comment on two versions of the *Administrative Effectiveness Survey*, each containing a different Likert scale heading (Appendix D). One panel consisted of 10 randomly selected faculty members from the University of Tennessee and the other panel consisted of a cluster of 20 doctoral students from the Educational Administration program at the University of Tennessee. Each version of the *Administrative Effectiveness Survey* had a different set of stem options for the response scale. One version of the questionnaire included a Likert scale ranging from “poor, fair, good, and excellent”. The second version of the questionnaire included a Likert scale ranging from “unsatisfactory, needs improvement, satisfactory, and exemplary”. The panels reviewed the two questionnaires and commented on the following:

- 1) Do you have a preference as to which response scale you would like to see on the questionnaire? Which response scale would you recommend?

- 2) Given the two versions of the questionnaire, would you rate the performance of your academic dean differently; that is, would you check a different response option for the same item on each questionnaire?

Responses from the two panels helped the researcher to determine which response scale was to be used during the pilot study as well as the subsequent national field study. From the 10 faculty members, only 3 recommended use of the scale “poor, fair, good, and excellent” and the other 7 recommended use of the scale “unsatisfactory, needs improvement, satisfactory, and exemplary”. All 20 of the doctoral students recommended use of the scale “unsatisfactory, needs improvement, satisfactory, and exemplary”. Very few people answered the second question. Two indicated that they would rate the performance differently based on the different response option and three indicated that they would not. Based on these results, the researcher chose to use the scale “unsatisfactory, needs improvement, satisfactory, and exemplary”.

One public and one private institution were selected from each of the four designated Carnegie classifications (based on *The 2000 Carnegie Classification of Institutions of Higher Education*: Doctoral/Research Universities-Extensive; Doctoral/Research Universities-Intensive; Master’s Colleges and Universities I; and Master’s Colleges and Universities II), for a total of eight institutions for the pilot study. The Chief Academic Officers at these institutions were contacted to see if they would agree to participate in this study. When one of the originally selected institutions did not agree to participate, another similar institution was contacted. After a process of contacting institutions several times, only four institutions agreed to participate in the pilot study, which was very disappointing to the researcher. The Chief Academic

Officers were asked to randomly select 10 faculty members to participate in the study. The data collection and treatment of data followed the method described below for the national field test.

The four participating institutions represented the four Carnegie classification types. One institution was private and the other three were public. One institution sent the names of 7 faculty members instead of 10 and so the sample size for the pilot study was set at 37 faculty members. Only 24 respondents completed the survey for a 65% response rate. While the response rate was fairly good, the sample size was not a large one.

SITE AND POPULATION

Four types of institutions were used in the national field test of the *Administrative Effectiveness Survey*, based on *The 2000 Carnegie Classification of Institutions of Higher Education*: Doctoral/Research Universities-Extensive; Doctoral/Research Universities-Intensive; Master's Colleges and Universities I; and Master's Colleges and Universities II. A systematic random selection was utilized to select one public and one private institution that had Colleges of Agriculture, Arts and Humanities, Business, Education, Engineering, Law, Nursing, and Social Work from six areas of the country, representing the six regional accrediting agencies, from each Carnegie classification.

Chief Academic Officers at each selected institution were contacted to solicit permission for faculty members of each of the colleges to participate in the study. When a CAO did not agree to allow his or her faculty members to participate in the study, another similar type of institution was selected and its CAO contacted. Four institutions

participated in the national study and a total of 677 faculty members completed the survey. The sample size for the study was 677.

DATA COLLECTION PROCEDURES

For each institution included in the sample, an individual packet requesting his or her participation in the research project was mailed to the Chief Academic Officer (CAO). The packet included: (1) a cover letter explaining who the researcher is as well as contact information and an institutional contact to address any participant or institutional questions, the purpose of the study, how the collected data will be used, and assurance that all information from the instrument will be confidential and only the researcher will have access to the data and any identifying information in order to meet the University of Tennessee Human Subjects Review Board requirements (Appendix E); (2) a short form to request the number of faculty members in each college, to provide a place for consent to participate in the study, and to ask the CAO to send a list of faculty names and e-mail addresses (Appendix F); and (3) a self-addressed stamped envelope.

Upon receipt from the CAO of his or her willingness to participate in the study, the CAO was contacted by e-mail and asked to notify his or her faculty of the study, to ask for faculty participation in the study, and to inform the faculty that within one week they will be receiving an e-mail with the URL of a web site location to access the evaluation instrument as well as a username and a unique password to access the site. The researcher then sent an e-mail to the faculty members participating in the study explaining the purpose of the study, how the collected data will be used, that all information from the instrument will be confidential, and that only the researcher will have access to the data and any identifying information (Appendix G). This e-mail to the

faculty members also included the URL of the web site where they could access the evaluation instrument, and a username and password to access the site. The faculty members could find the informed consent letter on the website and they could only access the instrument after accepting the terms of the informed consent letter (Appendix H). The expected return rate for this study was 50%.

Approximately two weeks after the initial e-mail to the faculty members, an e-mail was sent to thank participants that have completed the on-line survey and to remind non-respondents to complete the survey (Appendix I). The rate of return was still low so a second follow-up e-mail requesting their cooperation and stressing the importance of their response was sent. The return rate still did not meet the expectation for the study and a third e-mail was then sent about two weeks after the second e-mail. Only 179 faculty members completed the instrument out of 677, for a response rate of 26%.

A total of 121 institutions were contacted to participate in the national field test. Unfortunately, 40 declined to participate, 77 never responded, with an additional follow-up, and four institutions agreed to participate. Of these four institutions, three were public institutions and one was private, two were Master's I and two were Master's II. Data collection proved to be a much greater problem than was anticipated. The large number of schools which did not respond at all was not anticipated. In addition, known factors which contributed to the low institutional participation rate included institutions which required their own Institutional Review Board process for any research done on their campus, a number of other research projects that had already been conducted among faculty members at some campuses, institutions that had just completed their own internal evaluation system and did not want to participate in another one, and the

sensitive nature of the study subject which contributed to some institutional reluctance. The total number of faculty for the four institutions included in the national study was 677. The total number of respondents who completed the survey was 179, which yielded a response rate of 26%. There were 70 respondents from Master's I institutions, 44 were from Master's II, and 65 did not identify their Carnegie classification. This was much lower than the expected response rate of 50%. The data collection procedures and the design of this study received approval from the Institutional Research Board at the University of Tennessee.

TREATMENT OF THE DATA

In order to answer Research Question #1, "To what extent are the items seen as valid measures of administrator performance effectiveness as determined by two expert panels?" the questionnaire was sent to the two expert panels along with a cover page asking each panel member to return the instrument with his or her comments regarding the content and item construction of the instrument. A summary of the panel's suggestions was included in the write-up and was incorporated, if possible, into the instrument prior to the pilot study and the national field study.

The respondents completed the survey on-line and their responses were entered automatically into a Microsoft Access database. This data was then imported into a Microsoft Excel spreadsheet to facilitate data analysis. SPSS and SAS were used to analyze the survey responses.

Principal components analysis (PCA) was the next step in the treatment of the data. PCA is useful in determining the true dimensionality of the data set and can identify the number of factors to be used in a factor analysis (Johnson, 1998). Next,

factor analysis (FA) was conducted in order to answer Research Question #2, “To what extent does a factor analysis of the item responses reflect the initial scales of the instrument?” and to determine a new set of uncorrelated variables that better described the data being analyzed. Factor analysis can determine whether the evaluative items falls into clusters that constitute better factors to more accurately describe the data. For example, questions 6 through 13 of the survey instrument may actually represent two factors such as “Decision-making skills” and “Communication skills” instead of the descriptor “Leadership style and skill”.

In order to answer Research Question #2a, “Is there any difference in the evaluative factors that emerge across institution type when classified by Carnegie classification?” a factor analysis was to be conducted for each of the four Carnegie type classifications in order to determine whether the factor structure changed. However, as the results show in the next chapter, there were not enough institutions to participate in the national study to answer this question.

In order to answer Research Question #2b, “Is there any significant difference when the factor scores are compared to the overall rating score?” a multiple regression analysis was performed. This analysis determined if there were any correlations between the individual factor scores and the score from the overall evaluation item.

In order to answer Research Question #3, “What level of reliability is found for the Administrative Effectiveness Survey when the Cronbach Alpha coefficient is calculated?” and to establish reliability of the instrument, the Cronbach Alpha coefficient was calculated. The instrument used a Likert scale and so the Cronbach Alpha coefficient was more appropriate than the Kuder-Richardson coefficient. A reliability

coefficient of .70 or higher is deemed acceptable for scales used in research (Litwin, 1995) and served as the determining factor in whether the instrument reached an acceptable level or not. Other analyses were performed if deemed appropriate and applicable.

CHAPTER IV

ANALYSIS OF THE DATA

INTRODUCTION

This chapter presents the results of the data analyses completed on the final version of the *Administrative Effectiveness Survey* (Appendix J) instrument for both the pilot and the national studies. The instrument contained 34 evaluative items, organized into five subgroups, two questions asking if respondents would include open-ended questions for examples of performance, one item for the overall performance of the dean, and two open-ended questions regarding improvement to the instrument.

The respondents completed the instrument on-line, which created a database, and their responses were imported into a Microsoft Excel spreadsheet. SPSS and SAS were used to analyze the survey responses. The findings and results presented in this chapter are organized and discussed for both the pilot and the national studies as they relate to the three research questions that guided this study.

RESEARCH QUESTION #1

The first research question was, “To what extent are the items seen as valid measures of administrator performance effectiveness as determined by two expert panels?” The first expert panel asked to review the instrument consisted of four members of the UT faculty serving on the committee guiding this study. In general, they commented that the instrument was a good one. Their recommendations for changes to the instrument and which were implemented in the final version of the instrument (Appendix J) included:

- Removing the word “leadership” in the titles of the subcategories

- Revising the title of the last subcategory from “Leadership of the Dean” to “Academic Stewardship”
- Revising phrasing in question three to “educational vision”
- Revising wording of question four to “monitors progress towards”
- Revise wording in question 11 from “Effectively manages conflict” to “Manages conflict effectively”
- Moving background information to the end of the survey
- Adding “other” to list of colleges
- Revising question 21 to remove the word “community”

One member commented that the phrasing of the questions and the options did not appear to match very well and that the questions were not phrased in a neutral manner. In addition, this member did not feel that the response scale categories “unsatisfactory” and “needs improvement” were discretely different. “Unsatisfactory” was therefore changed to “seriously inadequate”. Two of the four members said they would include the open-ended questions, one did not comment on this area, and the fourth member would not include the open-ended questions. Additional questions were suggested to be added to the survey and included one about accessibility and one about being an advocate for the college, programs, faculty, and students.

The second expert panel consisted of national experts in their respective fields in higher education administration. The original list of proposed 9 members was expanded to 12. However, even with a follow-up letter, only 5 of those individuals responded. Three respondents recommended including the open-ended questions. One panel member recommended including the open-ended question, “Describe briefly one action, behavior, or situation in which the Dean exhibited commendable and exemplary leadership,” but not the question, “Describe briefly one action, behavior, or situation in which the Dean’s performance reflected a need for improvement,” because the respondent felt that the latter type of question might be “off-putting”. Another

respondent, however, suggested including the open-ended question, “Describe briefly one action, behavior, or situation in which the Dean’s performance reflected a need for improvement,” but not the question, “Describe briefly one action, behavior, or situation in which the Dean exhibited commendable and exemplary leadership.” One respondent also commented that he had used a similar dean’s administrative evaluation survey for the faculty at his institution and the results were “disastrous;” however, no further information was provided as to why the process was disastrous. A question about whether a formal evaluation process for the evaluation of the dean was suggested as well as one on whether the respondent would like to have this dean as his/her leader. Other suggestions for improvement to the survey which were incorporated into the final version included:

- Defining “effectiveness” in the direction statement “Please assess the effectiveness of the Dean of your academic college on the following items:”
- Adding a question about establishing goals for the college
- Revising question six to add “as needed”
- Moving question 17 up to the top of the category
- Adding a question on “Exhibit good judgment in personnel retention and improvement”
- Moving question 22 up in the category and revising to add “defines and”
- Revising question 23 to add “as appropriate”
- Moving question 27 up in the category and revising to add “and research”

After incorporating the suggestions made by the two expert panels, the final version of the *Administrative Effectiveness Survey* was utilized in the rest of the study (Appendix J). However, one must note that because only 5 of the 12 national expert panel members responded, the content validity of the AES was not established.

RESEARCH QUESTION #2

The second research question was, “To what extent does a factor analysis of the item responses reflect the initial scales of the instrument?” Twenty-four individuals completed the on-line survey during the pilot study. Some of those respondents, though, filled out “no basis for observation” on one or more of the questions and these responses were excluded from some of the data analyses. Pearson correlations were calculated between the overall evaluation item score and the five subcategories on the instrument to see if any correlation existed among the items. The Pearson correlations between the overall evaluation item score and planning, leadership, resources, values, and academics, were respectively: .851, .889, .822, .878, and .776. Significant correlation was found with a p-value of $<.01$ for the subcategories of Leadership Style and Skill (noted as Leadership in the analysis), Resource Stewardship (noted as Resources in the analysis), and Values and Standards (noted as Values in the analysis) and significant correlation at the .05 level was found for the subcategory of Academic Stewardship (noted as Academics in the analysis). Only the Planning and Goal Setting (noted as Planning in the analysis) subcategory was uncorrelated with the overall evaluation item (noted as “Ov” in the analysis). These results can be found in Table 1. The pilot study results indicated that many of the items in the instrument were highly correlated with one another and that there may not be very many factors.

For the national study, 179 respondents completed the on-line survey; however, some responses were excluded from portions of the data analysis because participants answered one or more questions with “No basis for observation”. Demographic information for the data from the national study sample can be found in Appendix L. A

Table 1**Pearson Correlations – Pilot Study**

		Ov	Planning	Leadership	Resources	Values	Academics	Total
Ov	Pearson Correlation	1	.353	.725**	.714**	.606**	.510*	.697**
	Sig. (2-tailed)		.099	.000	.000	.002	.013	.000
	N	23	23	23	23	23	23	23
Planning	Pearson Correlation	.353	1	.550**	.572**	.767**	.505*	.816**
	Sig. (2-tailed)	.099		.005	.004	.000	.014	.000
	N	23	24	24	23	24	23	24
Leadership	Pearson Correlation	.725**	.550**	1	.589**	.799**	.606**	.873**
	Sig. (2-tailed)	.000	.005		.003	.000	.002	.000
	N	23	24	24	23	24	23	24
Resources	Pearson Correlation	.714**	.572**	.589**	1	.609**	.692**	.800**
	Sig. (2-tailed)	.000	.004	.003		.002	.000	.000
	N	23	23	23	23	23	23	23
Values	Pearson Correlation	.606**	.767**	.799**	.609**	1	.653**	.929**
	Sig. (2-tailed)	.002	.000	.000	.002		.001	.000
	N	23	24	23	23	24	23	24
Academics	Pearson Correlation	.510*	.505*	.606**	.692**	.653**	1	.796**
	Sig. (2-tailed)	.013	.014	.002	.000	.001		.000
	N	23	23	23	23	23	23	23
Total	Pearson Correlation	.697**	.816**	.873**	.800**	.929**	.796**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	23	24	24	23	24	23	24

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

* . Correlation is significant at the 0.05 level (2-tailed).

principal component analysis revealed three factors (Table 2), although one factor emerged as the primary factor as it explained 65.05% of the variance. The primary factor was labeled as “Leadership” because the items that loaded on the factor related to planning, leadership, and values. The second factor contained items relating to external academic stewardship and was referred to as “External Academics”, and the third factor contained items regarding resources, supporting the mission of the college, and motivating colleagues to high standards and was designated as “Mission, Standards, and Resources”. Figure 1 provides a scree plot as a visual representation that there are only three factors and show the primary factor.

Table 3 contains the rotated component matrix which reveals what items load onto what factors. In order to create an instrument with the three factors (Leadership, External Academics, Mission/Standards/Resources) identified from this analysis and to be used for further analysis later, only items on the Rotated Component Matrix (Table 3) with a score of .575 or higher were kept, resulting in 28 items remaining in the instrument. Appendix K contains a copy of what the instrument would look like with these three factors and the items that represent each of them.

For the second research question, the results of the factor analysis indicated the need for only three scales instead of the original five. The three scales have been named “Leadership,” “External Academics,” and “Mission, Standards, and Resources”. Because of the limited number of responses from the faculty in the sample, the factor analysis analysis performed on the data may not be very dependable. Had more data been available, more factors might have emerged.

Table 2
Principal Component Analysis

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	22.117	65.049	65.049	22.117	65.049	65.049
2	2.206	6.487	71.536	2.206	6.487	71.536
3	1.181	3.473	75.008	1.181	3.473	75.008
4	.796	2.341	77.349			
5	.765	2.250	79.599			
6	.661	1.945	81.543			
7	.623	1.832	83.375			
8	.591	1.737	85.113			
9	.570	1.675	86.788			
10	.500	1.470	88.258			
11	.443	1.304	89.563			
12	.358	1.053	90.616			
13	.334	.981	91.597			
14	.313	.921	92.517			
15	.283	.832	93.350			
16	.273	.804	94.153			
17	.241	.710	94.863			
18	.220	.646	95.509			
19	.192	.566	96.075			
20	.179	.526	96.602			
21	.157	.462	97.064			
22	.141	.415	97.479			
23	.137	.404	97.883			
24	.115	.337	98.220			
25	.106	.312	98.532			
26	.102	.299	98.831			
27	.081	.237	99.068			
28	.074	.218	99.286			
29	.065	.190	99.476			
30	.053	.155	99.631			
31	.045	.133	99.765			
32	.037	.108	99.873			
33	.024	.072	99.944			
34	.019	.056	100.000			

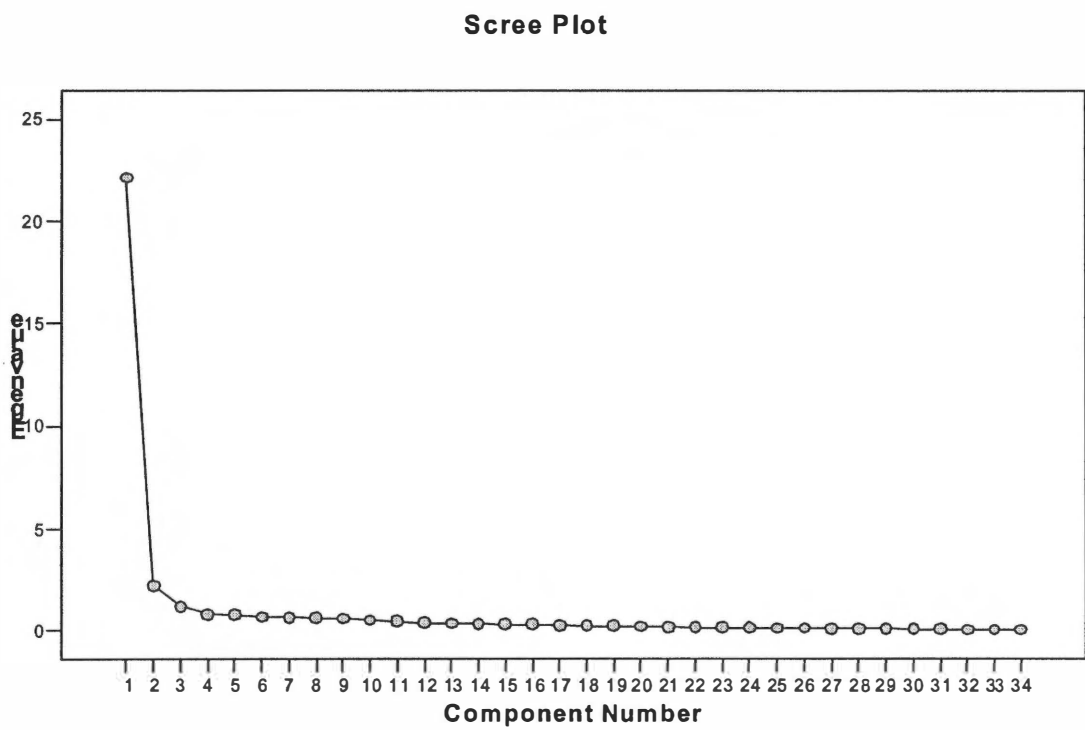


Figure 1. Scree Plot from Factor Analysis

Table 3**Rotated Component Matrix**

Item #	Component		
	1	2	3
1	.575		
2	.530	.558	.406
3	.520	.533	.406
4	.664	.424	
5	.622		
6	.455		.716
7	.823		
8	.835		
9	.747	.428	
10	.794		
11	.784		
12	.763		
13	.545		.531
14	.480	.444	.586
15	.654		.455
16	.663	.461	
17	.405		.630
18		.461	.716
19	.454	.418	.639
20		.458	.575
21	.807		
22	.771		
23	.704		.540
24		.420	.529
25	.746		.442
26	.785		
27	.710	.456	
28		.573	.470
29	.562	.564	
30		.805	
31		.851	
32		.716	
33		.747	.450
34		.638	.518

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 9 iterations.

RESEARCH QUESTION #2a

Research Question 2a was, “Is there any difference in the evaluation factors that emerge across institution type when classified by Carnegie classification?” The pilot study contained one institution from each of the four Carnegie Classification types: Doctoral/Research Universities – Extensive, Doctoral/Research Universities – Intensive, Master’s Colleges and Universities I, and Master’s Colleges and Universities II. This question could not be answered by a factor analysis for each of the four Carnegie classifications due to the small sample size. However, using the General Linear Model, multivariate tests were run to see if there was any significant difference between the factor scores and the four Carnegie types for the pilot study. The factor scores from the following factors, planning, leadership, resources, values, and academics, serve as the dependent variables to the independent variable, Carnegie. The p-values for each of the dependent variables and the independent variable, respectively, were .502, .161, .469, .203, .243, and .132. None of these were significant. Furthermore, Pillai’s Trace test had a p-value of .151 and Wilk’s Lambda test had a p-value of .206, both results supporting that the factor scores are not related to Carnegie classification type. Table 4 summarizes this information. However, there were so few respondents in the pilot study that these results are not very reliable.

In the national study, two of the institutions were Master’s I and two were Master’s II and no institutions from the other two Carnegie types participated in the study. Due to the fact that there were not respondents from all four Carnegie types, research question 2a could not be answered for the national study. This is an area for further research.

Table 4
Multivariate Tests – Pilot Study

Test of Between-Subjects Effect

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.
Carnegie	Ov	.604	3	.201	.814	.502
	Planning	1.277	3	.426	1.920	.161
	Leadership	.552	3	.184	.881	.469
	Resources	1.013	3	.338	1.687	.203
	Values	.703	3	.234	1.514	.243
	Academics	.615	3	.205	2.116	.132

Multivariate Analysis of Variance

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.992	303.781 ^a	6.000	14.000	.000
	Wilk's Lambda	.008	303.781 ^a	6.000	14.000	.000
	Hotelling's Trace	130.192	303.781 ^a	6.000	14.000	.000
	Roy's Largest Root	130.192	303.781 ^a	6.000	14.000	.000
Carnegie	Pillai's Trace	1.058	1.453	18.000	48.000	.151
	Wilk's Lambda	.260	1.358	18.000	40.083	.206
	Hotelling's Trace	1.762	1.240	18.000	38.000	.281
	Roy's Largest Root	.810	2.161 ^b	6.000	16.000	.102

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significant level.

RESEARCH QUESTION #2b

Research question 2b was, “Is there any significant difference when the factor scores are compared to the overall rating score?” For research question 2b, a multiple regression analysis was performed on the pilot study data. The regression method was utilized to calculate factor scores for the data by taking an adjusted average of the items that have high correlations with the factor. The five factor scores (from the subcategories) planning, leadership, resources, values, and academics were entered as the independent variables while the overall evaluation item was the dependent variable. The ANOVA had an F-value of 150.749, which is significant with a p-value of <0.01 and reveals a strong correlation between the factor scores and the overall evaluation score. Table 5 shows the model summary, the ANOVA, and the coefficient table that were calculated for the pilot study data.

For the national study, multiple regression statistics were calculated two different times. The first multiple regression was performed on the original five factors formed from the subcategories on the instrument. The independent variables were the factors planning, leadership, resources, values, leadership and the dependent variable was the overall evaluation item. The ANOVA resulted in an F-value of 150.749 and a significant p-value of <0.01 , indicating a strong correlation between the factor scores and the overall evaluation score (Table 6). The second multiple regression was performed on the three new factors – Leadership; External Academics; and Mission, Standards, and Resources – that emerged from the factor analysis and that only contained items with a value of .575 and above (from Table 3). Appendix K provides a copy of the *Administrative Effectiveness Survey* instrument with the items representing the three factors. The

Table 5**Multiple Regression – Pilot Study****Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the estimate
1	.838 ^a	.702	.614	.30511

a. Predictors: (Constant), academics, planning, leadership, resources, values

ANOVA^b

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	3.722	5	.744	7.996	.000 ^a
	Residual	1.583	17	.093		
	Total	5.304	22			

a. Predictors: (Constant), academics, planning, leadership, resources, values

b. Dependent Variable: Ov

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients		Sig.
		B	Std. Error	Beta	t	
1	(Constant)	.401	.644		.623	.541
	Planning	-.301	.206	-.306	-1.459	.163
	Leadership	.490	.240	.452	2.044	.057
	Resources	.635	.211	.605	3.005	.008
	Values	.285	.340	.236	.836	.415
	Academics	-.267	.297	-.182	-.899	.381

a. Dependent Variable: Ov

Table 6

Multiple Regression - National Study, Original Data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.911 ^a	.829	.824	.35322

a. Predictors: (Constant), academics, values, resources, planning, leadership

ANOVA^b

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	94.040	5	18.808	150.749	.000 ^a
	Residual	19.338	155	.125		
	Total	113.379	160			

a. Predictors: (Constant), academics, planning, leadership, resources, values

b. Dependent Variable: Ov

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients		Sig.
		B	Std. Error	Beta	t	
1	(Constant)	-.237	.150		-1.580	.116
	Planning	.146	.094	.125	1.548	.124
	Leadership	.455	.101	.437	4.528	.000
	Resources	.135	.091	.109	1.483	.140
	Values	.255	.116	.213	2.210	.029
	Academics	.093	.073	.074	1.262	.209

a. Dependent Variable: Ov

ANOVA resulted in an F-value of 100.009, which was significant with a p-value of <0.01. Table 7 contains the results of this regression analysis. The factor scores and the overall evaluation item were highly correlated.

RESEARCH QUESTION #3

The third research question was, “What level of reliability is found for the *Administrative Effectiveness Survey* when the Cronbach Alpha is calculated?” For the pilot study, reliability statistics were calculated and the Cronbach Alpha was .939 which is a very high level of reliability for the *Administrative Effectiveness Survey* instrument, given the available data (Table 8). Reliability statistics were also calculated for each of the five subcategories (Table 9). For the planning factor, the Cronbach Alpha was .869 and for the Leadership factor, the Cronbach Alpha was .858. The Cronbach Alpha for the Resource factor was .880 and for the Values factor, the Cronbach Alpha was .850. For the Academic factor, the Cronbach Alpha was .779. Although these alphas are quite high, the response rate for the pilot study was very low and the available data on which the analysis was conducted was low so the reliability of the AES based on this data is not very strong.

Reliability statistics were calculated from the data in the national study, and the Cronbach Alpha was .983. Again, this is a very high level of reliability for the *Administrative Effectiveness Survey* instrument (Table 10), but because of the low response rate in the study, one cannot necessarily assign that high level of reliability to the AES. Table 11 shows the calculated Cronbach Alpha for each subcategory from the instrument. The Cronbach Alpha for the Planning factor was .919 and for the Leadership factor, it was .957. For the Resource factor, the Cronbach Alpha was .913 and for the

Table 7**Multiple Regression – National Study, Three New Factors****Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.807 ^a	.651	.644	.50742

a. Predictors: (Constant), Leadership, External Academics, Mission/Standards/Resources

ANOVA^b

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	77.250	3	25.750	100.009	.000 ^a
	Residual	41.454	161	.257		
	Total	118.703	164			

a. Predictors: (Constant), Leadership, External Academics, Mission/Standards/Resources

b. Dependent Variable: Ov

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients		Sig.
		B	Std. Error	Beta	t	
1	(Constant)	.813	.157		5.164	.000
	Newfactor1	.051	.004	.834	11.449	.000
	Newfactor2	-.020	.010	-.106	-1.945	.054
	Newfactor3	.006	.015	.028	.398	.691

a. Dependent Variable: Ov

Table 8

Reliability Statistics – Pilot Study

Case Processing Summary

		N	%
Cases	Valid	9	37.5
	Excluded ^a	15	62.5
	Total	24	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.939	34

Table 9**Reliability Statistics for Subcategories for Pilot Study**Subcategory: Planning and Goal Setting

Case Processing Summary

		N	%
Cases	Valid	21	87.5
	Excluded ^a	3	12.5
	Total	24	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.869	6

Subcategory: Leadership Style and Skill

Case Processing Summary

		N	%
Cases	Valid	23	95.8
	Excluded ^a	1	4.2
	Total	24	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.858	8

Table 9. Continued.

Subcategory: Resource Stewardship

Case Processing Summary

		N	%
Cases	Valid	21	87.5
	Excluded ^a	3	12.5
	Total	24	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.880	5

Subcategory: Values and Standards

Case Processing Summary

		N	%
Cases	Valid	23	95.8
	Excluded ^a	1	4.2
	Total	24	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.850	8

Table 9. Continued.

Subcategory: Academic Stewardship

Case Processing Summary

		N	%
Cases	Valid	12	50.0
	Excluded ^a	12	50.0
	Total	24	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.779	7

Table 10

Reliability Statistics – National Study

Case Processing Summary

		N	%
Cases	Valid	70	39.1
	Excluded ^a	109	60.9
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.983	34

Table 11**Reliability Statistics for Subcategories for National Study**Subcategory: Planning and Goal Setting

Case Processing Summary

		N	%
Cases	Valid	113	63.1
	Excluded ^a	66	36.9
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.919	6

Subcategory: Leadership Style and Skill

Case Processing Summary

		N	%
Cases	Valid	134	74.9
	Excluded ^a	45	25.1
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.957	8

Table 11. Continued.

Subcategory: Resource Stewardship

Case Processing Summary

		N	%
Cases	Valid	136	76.0
	Excluded ^a	43	24.0
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.913	5

Subcategory: Values and Standards

Case Processing Summary

		N	%
Cases	Valid	131	73.2
	Excluded ^a	48	26.8
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.934	8

Table 11. Continued.

Subcategory: Academic Stewardship

Case Processing Summary

		N	%
Cases	Valid	79	44.1
	Excluded ^a	100	55.9
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.924	7

Values factor, the Cronbach Alpha was .934. The Cronbach Alpha for the Academic factor was .924. In addition, the Cronbach Alpha was calculated for each of the three new factors that were formed after the factor analysis was conducted (Table 12). For the Leadership factor, the Cronbach Alpha was .978 and for the External Academics factor, it was .898. The Cronbach Alpha for the Mission/Standards/Resources factor was .904. A Cronbach Alpha of .70 or higher is deemed as an acceptable level for establishing reliability (Litwin, 1995). However, as stated above, due to low response rate, this reliability statistic does not necessarily mean that the AES is a reliable instrument.

SUMMARY

Based on the data from the pilot study and the national study, tests of validity and reliability were conducted for the *Administrative Effectiveness Survey*, but these findings do not necessarily indicate that the validity or reliability of the instrument has been established. As will be discussed in the next chapter, due to data collection complications and lack of significant effects, these findings are not as substantial as the researcher had hoped to establish. Three expert panel reviews of the instrument and a factor analysis of the national study were conducted in an attempt to establish the content validity of the AES. The factor analysis revealed one primary factor and two secondary factors rather than the original five factors contained in the instrument. These three new factors were: Leadership, External Academics, and Mission, Standards, Resources. Reliability was analyzed using the Cronbach Alpha. The Cronbach Alpha for the original instrument was .983.

One other area of note relates to the two questions that were included on the instrument and which asked whether the respondents would like to see open-ended

Table 12**Reliability Statistics for New Factors for National Study****Leadership:****Case Processing Summary**

		N	%
Cases	Valid	104	58.1
	Excluded ^a	75	41.9
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.978	17

External Academics:**Case Processing Summary**

		N	%
Cases	Valid	82	45.8
	Excluded ^a	97	54.2
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.898	5

Table 12. Continued.

Mission, Standards, and Resources:

Case Processing Summary

		N	%
Cases	Valid	132	73.7
	Excluded ^a	47	26.3
	Total	179	100.0

(a). Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.904	5

questions in order to provide examples of exemplary performance or performance areas that need improvement. For the pilot study data, the majority of respondents would include both open-ended questions. For the first open-ended question regarding the inclusion of an example of exemplary performance, 71% would include it (Table 13). For the second open-ended question regarding the inclusion of an example of performance that needs to be improved, 63% would include it. The results were very similar for the data from the national study. For the first open-ended question, 72% would include such a question. For the second open-ended question, 67% would include such a question. Should this instrument be commercially marketed in the future, these results indicate that respondents would like the opportunity to complete open-ended questions and provide examples of performance to their academic deans.

Table 13
Responses Regarding Two Open-Ended Questions

Pilot Study

		N	%
Open-Ended #1	Yes	17	71
	No	7	29
	Total	24	100
Open-Ended #2	Yes	15	62.5
	No	9	37.5
	Total	24	100

National Study

		N	%
Open-Ended #1	Yes	126	72
	No	50	28
	Total	176	100
Open-Ended #2	Yes	118	67
	No	58	33
	Total	176	100

CHAPTER V

SUMMARY, FINDINGS, DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

SUMMARY

The purpose of this study was to revise and to establish the reliability and validity of the *Administrative Effectiveness Survey* and to field test the survey on the performance of academic deans at selected institutions of higher education. The survey was designed for faculty members to assess the job performance of their academic deans.

In conducting a study to establish a valid and reliable administrative performance instrument, a review of the literature revealed a paucity of available instruments, either institutional or commercial, that had established reliability and validity. Further, the review of literature revealed that many institutions are evaluating their academic deans but without any consistent method or a valid or reliable instrument. The initial instrument used in this study was revised after the review of literature, and additional questions were added regarding the performance of academic deans. The instrument was further revised when new response options were selected based on the results of two panels and their recommendations.

Content validity was to be established, in part, by sending the revised *Administrative Effectiveness Survey* to two expert panels for review. One panel consisted of the faculty guiding this study and the second panel contained national experts in their respective fields of higher education. These individuals made suggestions to revise the wording of some questions, to add two additional questions, and to revise some of the organization and formatting of the instrument. Most of these recommendations were

incorporated into the final version of the survey. This revised survey was not resubmitted to the expert panels for additional review, which could have added to a stronger claim of validity for this instrument. Factor analysis was performed on the data in order to attempt to further establish the validity of the items on instrument. The low response rate by the expert panel members and the respondents for the survey instrument hindered the establishment of content validity for the AES.

The survey was piloted at four institutions, one from each of the four different Carnegie type classifications – Doctoral/Research Universities-Extensive, Doctoral/Research Universities-Intensive, Master’s Colleges and Universities I, and Master’s Colleges and Universities II. Twenty-four respondents completed the on-line survey, which constituted a 65% response rate. The researcher conducted the statistical analysis of the data using SPSS and SAS programs. The purpose of the pilot study was to allow any necessary changes to be made to the instrument or the data collection method prior to the national field test of the instrument. No changes were made to the instrument after the pilot study; however, follow-up to gain institutional participation was changed to an electronic mail method in order to speed up the process. The data from the pilot study were not included in the analysis of the national study because that was not part of the original design methodology.

A systematic random selection method was initially employed to garner institutional participation. This portion of the study, though, proved to be most difficult because many Chief Academic Officers did not respond, even with follow-up e-mail; and a significant number declined. The researcher contacted 121 institutions and only 4 finally agreed to participate. The total number of faculty members asked to complete the

survey was 677, of which 179 completed the survey, representing a 26% response rate which was much lower than expected.

Factor analysis was performed on the data from the national study in order to further establish validity of the instrument and to determine the best number of evaluative scales for the instrument. Reliability statistics were calculated using the Cronbach Alpha coefficient in order to determine whether the instrument was reliable or not. The Cronbach Alpha for the *Administrative Effectiveness Survey* produced a reliability coefficient of .983, which is considered to be very high but does not mean that the AES is a reliable instrument. Results from asking whether two open-ended questions should be included on future instruments indicated that respondents would like to see such questions on the AES.

Data were coded and subjected to analysis through the SPSS and SAS programs. Pearson correlations, factor analysis, reliability statistics, multivariate tests, and multiple regression were used to determine specific results.

FINDINGS

Research Question #1: To what extent are the items seen as valid measures of administrator performance effectiveness as determined by two expert panels?

Feedback from most of the responding panel members indicated that the items were valid measures of administrator performance effectiveness. The recommendations made by most panel members related to wording of items, organization of the instrument, and adding two questions.

Research Question #2: To what extent does a factor analysis of the item responses reflect the initial scales of the instrument?

The factor analysis of the data from the national study revealed one primary factor, referred to as “Leadership,” that explained 65.05% of the variance and two secondary factors, “External Academics” and “Mission, Standards, and Resources”. The Leadership factor contained items mostly from the original subcategories of Planning and Goal Setting, Leadership Style and Skill, and Values and Standards. The second factor, labeled as External Academics, contained items related to external academic stewardship and the third factor, labeled as Mission, Standards, and Resources, contained an assortment of items regarding resources, college mission, and motivating colleagues to high performance standards.

The low response rate from the faculty, however, may have contributed to the fact that more factors were not found. While there was enough data to meet the statistical minimum for analysis, these findings may not be accurate and dependable.

Research Question #2a: Is there any variance in the evaluative factors that emerge across institution type when classified by Carnegie classification?

The national study did not contain an institution from each of the four Carnegie classification types and so this question could not be answered. However, results from the pilot study revealed no significant difference in variance of the evaluative factors across institution type when classified by Carnegie type (see Table 4).

Research Question #2b: Is there any significant difference when the factor scores are compared to the overall rating score?

There is not a significant difference when the factor scores are compared to the overall rating score. There is a strong correlation between the factor scores and the

overall evaluation item, which means that the factor scores predict how respondents will rate a dean on his or her overall performance.

Research Question #3: What level of reliability is found for the *Administrative Effectiveness Survey* when the Cronbach Alpha coefficient is calculated?

Reliability statistics were conducted for several different aspects of the survey. The Cronbach Alpha coefficient for the *Administrative Effectiveness Survey* was .983, which is a very high level of reliability. Cronbach Alphas were also calculated for each individual factor as well as for the three new factors that emerged from the factor analysis. The Cronbach Alphas for the individual factors, planning, leadership, resources, values, academics, were respectively: .919, .957, .913, .934, and .924. The Cronbach Alphas for the three new factors, leadership, external academics, mission/standards/resources, were respectively: .978, .898, and .904. All were well over .70, which is deemed as an acceptable level of reliability.

DISCUSSION

The researcher did not achieve the goals of this study, which was very disappointing. The task of the study was to revise and establish the validity and reliability of the *Administrative Effectiveness Survey*. The researcher experienced a great deal of frustration throughout the process and, in particular, in trying to secure institutional participation. Because of the low institutional response rate as well as the low response rate by the faculty members who completed the survey, the findings are tentative at best. This section will explain some of the difficulties that the researcher experienced and speculate on some ways that the survey design might have been strengthened.

Although the researcher made numerous attempts at encouraging institutional participation, this did not occur. From the first mailing, two institutions agreed to participate. From this point on, the challenge of getting more institutions to participate continued to grow, as did the researcher's disappointment and frustration level. After repeated additional mailings, at least three meetings with the advisor to this study, and calling some institutional contacts, the last mailing resulted in two additional institutions agreeing to participate. Most of the institutions that responded and declined to participate did not provide a reason for this decision. Only having four institutions participate in the national study was disheartening; however, the researcher continued with the data gathering portion of the study from the faculty at each of these institutions.

The researcher can speculate as to why some of the institutions did not participate, which might explain the low response rate, and can indicate ways that the methodology might have been strengthened. First, for those institutions that did not respond in any way, the Chief Academic Officers (CAO) may not have been interested in participating or have had someone else who opens his or her mail or e-mail and who may have deemed this study as not one that the university should undertake. Another possibility may have been that the CAO was out of the office when the packet or e-mail was received in the office and chose not to respond if he or she had missed the suggested deadline listed in the packet. The researcher was very frustrated with the lack of any kind of response because the researcher felt some kind of institutional response would have been professional.

Second, there were some institutions that required their own Institutional Review Board (IRB) process to be completed before any research could be conducted on the

campuses and did not have the personnel resources to complete this task for the researcher. The researcher communicated with one institutional representative who stated that the institution would have participated, but this person was unable to complete the IRB form to assist the researcher. A couple of institutions requested a copy of the IRB from the University of Tennessee and one agreed to participate after receiving that document. The design might have been strengthened at the beginning by including a copy of the University of Tennessee IRB form in the initial mailing or including a statement that the researcher would complete any institutional IRB form. A copy of the survey instrument could also have been included in the mailing.

Third, some schools indicated that they could not participate because the institution already had a number of research projects being conducted among the faculty and the Chief Academic Officer did not want to overload the faculty. A fourth reason was that some institutions had just conducted their own internal evaluation process and did not want to participate in another process for fear of confusing the faculty members. These are difficult reasons to overcome; however, better timing of the mailing might have contributed to a better design, as well as placing more emphasis on what the contribution of this study would be to the institution.

A fifth reason may have been the sensitive nature of the study. Asking people to evaluate job performance may cause some people to worry about a negative perception being placed upon themselves, the college, or the institution. More emphasis might have been placed on confidentiality of responses, that the results were only to be reported in aggregate form, and that institutions would not be identified. In addition, emphasis in the

follow-up e-mails might have been placed on the purpose of the study and that it was not to determine if the deans were doing a good or bad job.

Pressure from academic deans may have provided a sixth reason that some institutions did not participate. At one participating institution, the person coordinating institutional participation in the study encountered some pressure from the academic deans. After the researcher had sent out e-mails to individual faculty members to participate in the study, this person sent his own e-mail to the faculty. This person's e-mail mentioned some hesitancy in agreeing to allow institutional participation and stressed that the faculty did not have to participate in the study, even stating that the faculty could consider the survey another officious survey that institutions frequently receive. This may have contributed to a lower response rate from the faculty at this institution. In order to alleviate concern among some academic deans, the researcher might have asked the CAO if she could mail information to the academic deans informing them of the purpose of the study.

The response rate for the faculty members who actually completed the survey was also much less than the expected response rate for the study. The researcher was very diligent and timely in sending e-mails to the faculty and following up with those that had not completed the survey. The researcher actually followed up with those non-respondents twice to get additional participants. However, there are just some individuals who will not complete surveys. There were also some faculty members who did not feel qualified to complete the survey because they had not been at the university long enough and this reaction had not been expected. It might have been helpful to have had questions regarding length of service and how often the faculty members interacted with the dean in

order to determine whether the faculty member could provide qualified responses to the instrument.

The low rate of institutional participation hindered the study in several ways. First, there were fewer available faculty members who could be asked to complete the survey instrument. A larger sample may have resulted in additional responses which would have strengthened the results of the statistical analyses. Next, there were no institutions that participated from all four Carnegie classifications and so one of the research questions could not be answered. Third, due to the low faculty response rate, the statistical analyses performed on the data were not as dependable as the researcher had hoped. The factor analysis revealed one primary factor, leadership, and two secondary factors, external academics and mission, resources, and standards. If there had been more data available, the factor analysis may have resulted in additional or different factors. The evaluative factors that suggest a set of criteria for evaluating academic deans may not have suggested the best set of criteria. Further, the reliability statistics were quite high for the AES, but one cannot say that the instrument is a reliable one since the response rate was so low. Thus, because of the low response rate, the AES cannot be established as an instrument with validity from the factor analysis or with reliability.

Another area for discussion is the lack of response from the national expert panel members. Because there were only 5 of the 12 national expert panel members who reviewed the instrument and provided feedback on its validity, content validity of the AES cannot be established. Additional experts need to be identified and contacted, perhaps in person, in order to establish the content validity of the AES. Having only five

responses does not provide a breadth of analysis or support for any claims of content validity.

Another method of garnering institutional participation might have been more productive. One such method could have been to narrow the sample of institutions and to schedule meetings with the Chief Academic Officers at those institutions in order to personally ask for their participation. The CAO might find it more difficult to deny participation in person to the researcher.

The difficulties encountered by the researcher throughout the data collection were very disheartening. The validity and reliability that were attained were not substantial. Before this instrument could be made commercially available, additional attempts need to be made in order to establish the validity and reliability of the instrument.

CONCLUSIONS

The following section presents conclusions drawn from the investigation of this study.

1. Although two expert panels reviewed the items on the *Administrative Effectiveness Survey*, content validity has not been established. Thus, the *Administrative Effectiveness Survey* may not measure what it purports to measure.
2. The reliability of the *Administrative Effectiveness Survey* has also not been fully established, even though the reported Cronbach Alpha was .983. Thus, the AES might not be very reliable when used across multiple testing sites.
3. Additional tests of validity and reliability will have to be conducted with the *Administrative Effectiveness Survey* and with larger response rates in order to implement its use as a commercial instrument.

4. The two-opened questions should be included on the *Administrative Effectiveness Survey*. Respondents want an opportunity to provide open ended feedback to their academic deans on specific areas of performance.

RECOMMENDATIONS

Data, findings, and conclusions presented in this study show that the *Administrative Effectiveness Survey* was revised and established with a limited measure of validity and reliability; however, the results were nowhere near what the researcher had hoped to achieve. Recommendations for further study are given below.

1. The study was delimited to include only four Carnegie classification type institutions. However, only two types are represented in the national study. Similar studies should be done with institutions from the other Carnegie classifications.
2. A limitation of the study was that the sample might not be racially diverse and that was the case in this study (see Appendix L). Similar studies at more diverse institutions might result in a more diverse sample.
3. An assumption of the study was that the faculty would be qualified to provide an evaluation of their academic dean. However, if a faculty member had not been at an institution very long or did not have much interaction with the dean, then this was difficult. Questions regarding length of service and amount of interaction may be useful to add and then a new study should be conducted.
4. Further validity studies with a wider range of expert panel members and a greater response from those members must be done in order to establish the content validity of the instrument with any sense of credibility. A study could be

designed whereby a benchmark was set for a certain percentage of panel members to agree that each item should be included on the instrument in order for the item to be retained.

5. The evaluative items on this instrument serve as only one set of criteria to measure the effectiveness of the performance of academic deans. Additional items may also contribute to that goal and still maintain the reliability of the instrument. Appendix M contains a list of suggested questions from the respondents to be added to the instrument or to be revised. Further research could include a review of those items and a subsequent determination as to whether to make any revisions or not. Appendix N also contains a list of suggestions from the respondents to improve the instrument. Some of those may very well need to be considered. However a review by experts should be conducted if any changes are made to the instrument.
6. A study could be conducted to compare the *Administrative Effectiveness Survey* with another instrument and to see which one has greater validity and reliability.

LIST OF REFERENCES

LIST OF REFERENCES

- Administrator Evaluation. (2000, July). Retrieved from <http://www.policy.ilstu.edu/>
- Aleamoni, L. M. (1995). Review of the Employee Effectiveness Profile. In J. C. Conoley & J. C. Impara (Eds), *12th Mental Measurements Yearbook*, (pp. 346-347). Lincoln, NE: The University of Nebraska Press.
- Astin, A. W. (1993). *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*. Phoenix: The American Council on Education and The Oryx Press.
- Austin, M. J., Ahearn, F. L., English, R. A. (Eds.). (1997). "Editor's Notes." *New Directions for Higher Education*, 25(98), 1-8.
- Barrett, M. A. (1993). "Issues and Problems of Performance Appraisal Systems for University General Staff: A Study in a U. S. University." *Journal of Tertiary Education Administration*, 15(1), 157-179.
- Belcher, M. J. (1992). *A Performance Feedback System for Administrators in Higher Education*. Paper presented at the Annual Meeting of the American Evaluation Association (Seattle, WA, November 5-7). (ERIC Document Reproduction Service No. ED 354265).
- Berdahl, R. O. & McConnell, T. R. (1999). "Autonomy and Accountability: Who Controls Academe?" In Altbach, P. G., Berdahl, R. O., & Gumport, P. J. (Eds.), *American Higher Education in the Twenty-first Century*, (pp. 70-88). Baltimore, MD: The Johns Hopkins University Press.

Birnbaum, R. (1992). *How Academic Leadership Works: Understanding Success and Failure in the College Presidency*. San Francisco: Jossey-Bass, Inc.

Bogue, E. G. & Aper, J. (2000). *Exploring the Heritage of American Higher Education: The Evolution of Philosophy and Policy*. Phoenix: Oryx Press.

Bogue, E. G. & Hall, K. B. (2003). *Quality and Accountability in Higher Education: Improving Policy, Enhancing Performance*. Westport, CT: Praeger Publishers.

Bogue, E. G. & Noland, B. (2003). *The Evolution and Effectiveness of Accountability Policy in Higher Education: A Multi State Comparative Case Study*. Unpublished policy research and development proposal.

Bogue, E. G. & Saunders, R. L. (1992). *The Evidence for Quality: Strengthening the Tests of Academic and Administrative Effectiveness*. San Francisco: Jossey-Bass.

Bok, D. C. (1986). *Higher Learning*. The President and Fellows of Harvard College.

Bourque, L. B. & Fielder, E. P. (1995). *How To Conduct Self-Administered And Mail Surveys*. Thousand Oaks, CA: SAGE Publications.

Bowker, L. H. (1982). "The Academic Dean: A Descriptive Study." *Teaching Sociology*, 9(3), 257-271.

Boyer, E. L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate*. The Carnegie Foundation for the Advancement of Teaching.

Brown, D. E. & Rodriguez, R. C. (1989). "A Valid Self-Appraisal Instrument for Community College Administrators." *Community Junior College Quarterly of Research and Practice*, 13(3-4), 149-156.

Budig, J. E. (1995). *A Model for Administrator Evaluation by Subordinates*. Paper presented at the Annual Forum of the Association for Institutional Research (Boston, MA, May 28-31). (ERIC Document Reproduction Service No. ED 385325).

Burd, S. (2002). "Accountability or Meddling?" *The Chronicle of Higher Education*, September 20, 2002, pp. A23-A25.

Burke, J., & Seban, A. (Eds.). (1998). "Performance Funding for Higher Education: Fad or Trend?" [Special issue]. *New Directions for Institutional Research*, No. 97, Spring.

Cahn, S. M. (1994). *Saints and Scamps: Ethics in Academia*. New Jersey: Rowman & Littlefield.

Callan, P. M. (1993). "Government and Higher Education." In A. Levine (Ed.), *Higher Learning in America 1980-2000* (pp. 3-19). Baltimore: The Johns Hopkins University Press.

Dill, W. R. (1980). "The Deanship: An Unstable Craft." In Griffiths, D. E. & McCarty, D. J. (Eds.). *The Dilemma of the Deanship*. Danville, IL: The Interstate Printers and Publishers, Inc. 261-284.

Dipboye, R. L. (1978). Review of the Leader Behavior Description Questionnaire. In O. K. Buros (Ed.), *8th Mental Measurements Yearbook* (pp. 1745-1747). Highland Park, NJ: The Gryphon Press.

Enarson, H. (1962). "The Academic Vice-President or Dean." In Burns, G. P. (Ed.). *Administrators in Higher Education: Their Functions and Coordination*. New York: Harper & Brothers. 111-124.

Ewell, P. T. (1994). "A Matter of Integrity: Accountability and the Future of Self-Regulation." *Change*, 26(6), 25-29.

Ewell, P. T., Finney, J., & Smart, J. (1990). "Filling in the Mosaic: The Emerging Pattern of State-Based Assessment." *AAHE Bulletin*, 42 (8), 3-5.

Frank, M.J. (1994, March 28). Faculty moving forward with evaluation of deans. Retrieved April 12, 2003 from <http://www.umich.edu/>

Fox, D. J. (1969). *The Research Process in Education*. New York: Holt, Rinehart, and Winston, Inc.

Fujita, E. (1990). *The Evaluation of College Presidents: Dimensions Used by Campus Leaders*. Paper presented at the Annual Meeting of the Association for the Study of Higher Education (Portland, OR, November 1-4). (ERIC Document Reproduction Service No. ED 326114).

Gaither, G., Nedwek, B. P., & Neal, J. E. (1994). *Measuring Up: The Promises and Pitfalls of Performance Indicators in Higher Education*. ASHE-ERIC Higher Education Report No. 5. Washington, D.C.: The George Washington University, Graduate School of Education and Human Development.

Goldman, B. A. & Mitchell, D. F. (2003). *Directory of Unpublished Experimental Mental Measures*, vol. 8. Washington, D.C.: American Psychological Association.

Goldman, B. A., Mitchell, D. F., Egelson, P. E. (1977). *Directory of Unpublished Experimental Mental Measures*, vol. 7. Washington, D.C.: American Psychological Association.

Halpin, A. W. (1957). *Manual for the Leader Behavior Description Questionnaire*. Columbus, OH: Fisher College of Business, The Ohio State University.

Harada, M. (1991). *Criteria for Evaluation of Community College Instructional Deans*. (ERIC Document Reproduction Service No. ED 338293).

Harcleroad, F. F. & Eaton, J. S. (2005). "The Hidden Hand: External Constituencies and Their Impact." In Altbach, P. G., Berdahl, R. O., & Gumport, P. J. (Eds.), *American Higher Education in the Twenty-first Century*, (pp. 253-283). Baltimore, MD: The Johns Hopkins University Press.

Hoyt, D. P. & Spangler, R. K. (1978). Administrative Effectiveness of the Academic Department Head: Correlates of Effectiveness. Research Report No. 42. Manhattan: Office of Educational Research, Kansas State University.

Johnson, D. E. (1998). *Applied Multivariate Methods for Data Analysis*. California: Brooks/Cole Publishing Company.

Keig, L. & Waggoner, M. D. (1994). *Collaborative Peer Review: The Role of Faculty in Improving College Teaching*. ASHE-ERIC Higher Education Report No. 2. Washington, D.C.: The George Washington University, School of Education and Human Development.

Kerr, C. (1994). "Forward: American Society Turns More Assertive: A New Century Approaches for Higher Education in the United States." In Altbach, P. G.,

Berdahl, R. O., & Gumpert, P. J. (Eds.). *Higher Education in American Society* (3rd ed.,) (pp. i-ix), Amherst: Prometheus Books.

Knight, W. H. & Holen, M. C. (1985). "Leadership and the Perceived Effectiveness of Department Chairpersons." *Journal of Higher Education*, 56(6), 677-690.

Lahti, R. E. (1978). "Managerial Performance and Appraisal." *New Directions for Higher Education*, 6(2), 1-10.

Lasley, T. J. & Haberman, M. (1987). "How Do University Administrators Evaluate Education Deans?" *Journal of Teacher Education*, 38(5), 13-16.

Lefton, R. E., Buzzota, V. R., Sherberg, M. & Karraker, D. L. (1977). *Effective Motivation Through Performance Appraisal*. New York: John Wiley & Sons.

Licata, C. M. (1999). "Precepts for Post-Tenure Reviews." *Trusteeship*, 7(6), 8-13.

Litwin, M. S. (1995). *How to Measure Reliability and Validity*. Thousand Oaks, CA: SAGE Publications.

Lovell, C. D. (2000). "Past and Future Pressures and Issues of Higher Education: State Perspectives." In J. Losco and B. L. Fife (Eds.), *Higher Education in Transition: The Challenges of the New Millennium* (pp. 109-131). Westport, CT: Bergin & Garvey.

Lynch, D. M. & Bowker, L. H. (1984). *Graduate Deans and Graduate Education: A National Study*. Indiana University of Pennsylvania. Institute for Advanced Research. (ERIC Document Reproduction Service No. ED 247872).

McKeachie, W. J. & Others. (1971). "Student Ratings of Teacher Effectiveness." *American Educational Research Journal*, 8, 435-445.

McKinnon, N. C. (1993). *Development of a Performance Appraisal Program for Nonacademic Staff at Atlantic Baptist College*. (ERIC Document Reproduction Service No. ED 358799).

Martin, J. L. (1993). *Academic Deans: An Analysis of Effective Academic Leadership at Research Universities*. Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April). (ERIC Document Reproduction Service No. ED 362079).

Matczynski, T., Lasley, T. J., & Haberman, M. (1989). "The Deanship: How Faculty Evaluate Performance." *Journal of Teacher Education*, 40(6), 10-14.

Miller, R. I. (1972). *Evaluating Faculty Performance*. San Francisco, CA.: Jossey-Bass, Inc.

Miller, R. & Miller, P. (2000). "Evaluating Collegiate Administrators." In Hoffman, A. M. & Summers, R. W. (Eds.). *Managing Colleges and Universities: Issues for Leadership*. Westport, CT: Bergin & Garvey. 199-209.

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.

Montez, J. M. (2002). Developing and piloting the Higher Education Leadership Instrument (HELI): "Bootstrapping" theory and measurement (Doctoral dissertation, Washington State University, 2002). *Dissertation Abstracts International*, 63, 07A.

Morris, V. C. (1981). *Deaning, Middle Management in Academe*. Urbana, IL: University of Illinois Press.

- Murphy, J. L., Impara, J. C., & Plake, B. S. (Eds). (1999). *Tests in Print V*. Lincoln, NE: Buros Institute of Mental Measurements.
- Ohmann, R. (2000). "Historical Reflections on Accountability." *Academe*, January-February, 24-29.
- Ory, J. C. (2000). "Managing Evaluations in Higher Education." In Hoffman, A. M. & Summers, R. W. (Eds). *Managing Colleges and Universities: Issues for Leadership*. Westport, CT: Bergin & Garvey. 187-198.
- Peters, R. (1994). "Accountability and the End(s) of Higher Education." *Change*, 26(6), November/December, 16-23.
- Pinero, A. C. (1988). Development and Validation of an Instrument to Measure Expectations of the Role of the Academic Dean in Puerto Rico (Doctoral dissertation, The Pennsylvania State University, 1988). *Dissertation Abstracts International*, 49, 07A.
- Rasmussen, G. R. (1978). "Evaluating the Academic Dean." *New Directions for Higher Education*, 6(2), 23-40.
- Remmers, H. H. & Weisbrodt, J. A. (1965). *Manual of Instructions for the Purdue Rating Scale for Instruction*. West Lafayette, IN: University Book Store.
- Robillard, D., Jr. (2000). "Toward a Definition of Deaning." *New Directions for Community Colleges*, 28(109), 3-8.
- Robles, H. J. (1998). *Reform in American Public Higher Education*. National Commission on Excellence in Education. (ERIC Document Reproduction Service No. ED 429520).

Sarros, J. C., Gmelch, W. H., & Tanewski, G. A. (1998). "The Academic Dean: a Position in Need of a Compass and Clock." *Higher Education Research and Development*, 17(1), 65-88.

Scott, R. A. (1979). "The 'Amateur Dean' in a Complex University: An Essay on Role Ambiguity." *Liberal Education*, 65(4), 445-452.

Skipper, C. E. (1977). "Administrative Skills of Effective and Ineffective University Leaders." *College and University*, 52, 276-279.

Skipper, C. E. (1982). *Four Indicators of Administrative Effectiveness*. Paper presented at the American Educational Research Association (New York, NY, March 22). (ERIC Document Reproduction Service No. ED 216634).

Skipper, C. E. (1976). "Personal Characteristics of Effective and Ineffective University Leaders." *College and University*, 51, 138-141.

Stronge, J. H. & Helm, V. M. (1991). *Evaluating Professional Support Personnel in Education*. Newbury Park, CA: SAGE Publications, Inc.

Surwill, B. J. & Heywood, S. J. (1976). *Evaluation of College and University Top Brass: The State of the Art, Status Report of AACSU Member Institutions*. (ERIC Document Reproduction Service No. ED 129142).

Texas A & M University-Commerce. (1996, September 1). Academic administrator evaluation. Retrieved April 12, 2003 from <http://www.tamu-commerce/edu/>

UMKC Faculty Senate. (1996, October 29). Evaluations of administrators. Retrieved April 12, 2003 from <http://www.umkc.edu/>

Wattenbarger, J. L. (1983). "Evaluation of College Administrators." In Smith, A. (Ed). "Evaluating Faculty and Staff". *New Directions for Community Colleges*, 41, 45-55.

Whitman, N. & Weiss, E. (1982). *Faculty Evaluation: The Use of Explicit Criteria for Promotion, Retention, and Tenure*. AAHE-ERIC Higher Education Report No. 2. Washington, D.C.: American Association for Higher Education.

Wilson, R. (1998). "New Research Casts Doubt on Value of Student Evaluations of Professors." *The Chronicle of Higher Education*, 44(19), p. A14.

Wilson, R. A. & Kroesser, H. L. (1981, Spring). "Administrative Evaluation: A Comprehensive Approach." *Journal of the College and University Personnel Association*, 32(1), 42-29.

Wolverton, M., Montez, J., & Gmelch, W. H. (2000). *The Roles and Challenges of Deans*. Paper presented at the Annual Meeting of the Association for the Study of Higher Education (Sacramento, CA, November 16-19). (ERIC Document Reproduction Service No. ED 449743).

Yang, S. (1987). The Development of an Appraisal Instrument for the Evaluation of Department Chairpersons in Taiwan Private Junior Colleges of Engineering and/or Business (Doctoral dissertation, University of Hawaii, 1987). *Dissertation Abstracts International*, 48, 11A.

APPENDICES

APPENDIX A

FACULTY SENATE ADMINISTRATIVE QUESTIONNAIRE

FACULTY SENATE ADMINISTRATIVE EVALUATION QUESTIONNAIRE
(Administrative Position to be entered here)

Please check the appropriate box.

Position:	☐ President's Office/ Staff ☐ Provost/Vice President of Academic Affairs/ Staff ☐ Other Vice Chancellor/Staff ☐ Dean/Dean's Office Staff	☐ Department/Unit Chair ☐ Member of Faculty Senate or Faculty Senate Committee ☐ Faculty ☐ Other (please specify): _____
Gender:	☐ Female ☐ Male	
Primary Race:	☐ African American ☐ Asian ☐ Hispanic ☐ American Indian ☐ Caucasian ☐ Pacific Islander	

	Strongly Disagree	Disagree	Agree	Strongly Agree	No Basis for Evaluation
LEADERSHIP PLANNING AND GOAL SETTING					
1. Establishes appropriate goals for self and office.	☐	☐	☐	☐	☐
2. Expects others to establish appropriate goals.	☐	☐	☐	☐	☐
3. Encourages creation of a shared vision.	☐	☐	☐	☐	☐
4. Monitors progress and performance of plans/goals and expects others to do the same.	☐	☐	☐	☐	☐
5. Respects the university's mission and reflects appropriate balance in emphasis.	☐	☐	☐	☐	☐
LEADERSHIP STYLE AND SKILL					
6. Takes responsibility for decisions and adapts decision style to situation.	☐	☐	☐	☐	☐
7. Empowers and encourages others to take responsibility for decisions.	☐	☐	☐	☐	☐
8. Communicates effectively and shares information on activity/achievement.	☐	☐	☐	☐	☐
9. Balances consensus and conviction in decision-making.	☐	☐	☐	☐	☐
10. Responds effectively to complaints and criticism.	☐	☐	☐	☐	☐
11. Manages conflict effectively and respects honest dissent.	☐	☐	☐	☐	☐
12. Represents office and university with poise and credibility.	☐	☐	☐	☐	☐
13. Inspires and motivates colleagues to high performance standards.	☐	☐	☐	☐	☐
LEADERSHIP RESOURCE STEWARDSHIP					
14. Is an effective steward of resources entrusted to the care of the office/university.	☐	☐	☐	☐	☐
15. Expects and encourages others to be effective resource stewards.	☐	☐	☐	☐	☐
16. Effectively prioritizes use of resources.	☐	☐	☐	☐	☐
17. Exhibits good judgment in personnel selection and evaluation.	☐	☐	☐	☐	☐

APPENDIX B

ADMINISTRATIVE EFFECTIVENESS SURVEY – ONE
ADMINISTRATIVE EFFECTIVENESS SURVEY – TWO

ADMINISTRATIVE EFFECTIVENESS SURVEY – ONE

Please check the appropriate box:

Gender: ☐ Female ☐ Male Primary Race: ☐ African American ☐ Asian ☐ Hispanic ☐ Caucasian ☐ Pacific Islander ☐ Other	College: ☐ Agriculture ☐ Arts and Humanities ☐ Business ☐ Education ☐ Engineering ☐ Law ☐ Nursing ☐ Social Work
--	--

Please assess the effectiveness of the Dean of your academic college on the following items:

	Unsatisfactory	Needs Improvement	Satisfactory	Exemplary	No Basis for Evaluation
LEADERSHIP PLANNING & GOAL SETTING					
1. Establishes appropriate goals for self.	☐	☐	☐	☐	☐
2. Assists others in establishing appropriate goals.	☐	☐	☐	☐	☐
3. Encourages creation of a shared vision.	☐	☐	☐	☐	☐
4. Monitors performance of plans/goals.	☐	☐	☐	☐	☐
5. Respects the university's mission.	☐	☐	☐	☐	☐
LEADERSHIP STYLE AND SKILL					
6. Adapts decision style to situation.	☐	☐	☐	☐	☐
7. Empowers others to take responsibility for decisions.	☐	☐	☐	☐	☐
8. Communicates effectively.	☐	☐	☐	☐	☐
9. Balances consensus and conviction in decision-making.	☐	☐	☐	☐	☐
10. Responds effectively to complaints.	☐	☐	☐	☐	☐
11. Effectively manages conflict.	☐	☐	☐	☐	☐
12. Represents the college with credibility.	☐	☐	☐	☐	☐
13. Motivates colleagues to high performance standards.	☐	☐	☐	☐	☐

LEADERSHIP RESOURCE STEWARDSHIP

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 14. Effectively manages resources entrusted to the care of the office/university. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 15. Encourages others to be effective resource stewards. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 16. Effectively prioritizes use of resources. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 17. Encourages others to be effective resource stewards. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 18. Effectively prioritizes use of resources. | ☐ | ☐ | ☐ | ☐ | ☐ |

LEADERSHIP VALUES AND STANDARDS

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 19. Treats colleagues with respect. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 20. Exhibits an appropriate balance of compassion and courage in relationships. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 21. Reflects personal integrity. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 22. Emphasizes both diversity and community in university life. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 23. Expects high standards of performance in programs. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 24. Builds a climate of trust in policy formation. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 25. Challenges ineffective practices. | ☐ | ☐ | ☐ | ☐ | ☐ |

LEADERSHIP OF THE DEAN

- | | | | | | |
|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 26. Effectively represents faculty issues to university administration. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 27. Fosters excellence in teaching methods among faculty. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 28. Effectively builds relationships with external stakeholders. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 29. Serves as an effective external raiser of funds for the college. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 30. Effectively models scholarship by publishing and/or presenting papers regularly. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 31. Encourages faculty involvement in community projects and endeavors. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 32. Emphasizes scholarly development and research among faculty members. | ☐ | ☐ | ☐ | ☐ | ☐ |

LEADERSHIP EVALUATION

- | | | |
|---------------------|---------------------------------------|--|
| Overall Evaluation: | ☐ Exemplary | ☐ Needs Improvement |
| | ☐ Satisfactory | ☐ Unsatisfactory |
-

If you were to use this instrument on your campus, would you include the following open-ended questions?

- ☐ Yes 1) Describe briefly one action, behavior, or situation in which the Dean exhibited commendable and exemplary leadership.
- ☐ No

- ☐ Yes 2) Describe briefly one action, behavior, or situation in which the Dean's performance reflected a need for improvement.
- ☐ No

What other questions would you include in this instrument?

What recommendations would you make for improving this evaluation instrument.

ADMINISTRATIVE EFFECTIVENESS SURVEY – TWO

Please check the appropriate box:

Gender: ☐ Female ☐ Male Primary Race: ☐ African American ☐ Asian ☐ Hispanic ☐ Caucasian ☐ Pacific Islander ☐ Other	College: ☐ Agriculture ☐ Arts and Humanities ☐ Business ☐ Education ☐ Engineering ☐ Law ☐ Nursing ☐ Social Work
--	--

Please assess the effectiveness of the Dean of your academic college on the following items:

	Poor	Fair	Good	Excellent	No Basis for Evaluation
LEADERSHIP PLANNING & GOAL SETTING					
1. Establishes appropriate goals for self.	☐	☐	☐	☐	☐
2. Assists others in establishing appropriate goals.	☐	☐	☐	☐	☐
3. Encourages creation of a shared vision.	☐	☐	☐	☐	☐
4. Monitors performance of plans/goals.	☐	☐	☐	☐	☐
5. Respects the university's mission.	☐	☐	☐	☐	☐
LEADERSHIP STYLE AND SKILL					
6. Adapts decision style to situation.	☐	☐	☐	☐	☐
7. Empowers others to take responsibility for decisions.	☐	☐	☐	☐	☐
8. Communicates effectively.	☐	☐	☐	☐	☐
9. Balances consensus and conviction in decision-making.	☐	☐	☐	☐	☐
10. Responds effectively to complaints.	☐	☐	☐	☐	☐
11. Effectively manages conflict.	☐	☐	☐	☐	☐
12. Represents the college with credibility.	☐	☐	☐	☐	☐
13. Motivates colleagues to high performance standards.	☐	☐	☐	☐	☐

LEADERSHIP RESOURCE STEWARDSHIP

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 14. Effectively manages resources entrusted to the care of the office/university. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 15. Encourages others to be effective resource stewards. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 16. Effectively prioritizes use of resources. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 17. Encourages others to be effective resource stewards. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 18. Effectively prioritizes use of resources. | ☐ | ☐ | ☐ | ☐ | ☐ |

LEADERSHIP VALUES AND STANDARDS

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 19. Treats colleagues with respect. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 20. Exhibits an appropriate balance of compassion and courage in relationships. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 21. Reflects personal integrity. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 22. Emphasizes both diversity and community in university life. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 23. Expects high standards of performance in programs. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 24. Builds a climate of trust in policy formation. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 25. Challenges ineffective practices. | ☐ | ☐ | ☐ | ☐ | ☐ |

LEADERSHIP OF THE DEAN

- | | | | | | |
|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 26. Effectively represents faculty issues to university administration. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 27. Fosters excellence in teaching methods among faculty. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 28. Effectively builds relationships with external stakeholders. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 29. Serves as an effective external raiser of funds for the college. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 30. Effectively models scholarship by publishing and/or presenting papers regularly. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 31. Encourages faculty involvement in community projects and endeavors. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 32. Emphasizes scholarly development and research among faculty members. | ☐ | ☐ | ☐ | ☐ | ☐ |

LEADERSHIP EVALUATION

- | | | |
|---------------------|---------------------------------------|--|
| Overall Evaluation: | ☐ Exemplary | ☐ Needs Improvement |
| | ☐ Satisfactory | ☐ Unsatisfactory |

If you were to use this instrument on your campus, would you include the following open-ended questions?

- ☐ Yes 1) Describe briefly one action, behavior, or situation in which the Dean exhibited commendable and exemplary leadership.
- ☐ No

- ☐ Yes 2) Describe briefly one action, behavior, or situation in which the Dean's performance reflected a need for improvement.
- ☐ No

What other questions would you include in this instrument?

What recommendations would you make for improving this evaluation instrument.

APPENDIX C

LETTER TO MEMBERS OF THE EXPERT PANELS

[Letterhead]

[Date]

[Name]

[Address]

Dear [Name]:

With the contemporary accent on accountability in American higher education, appraisal of both faculty and administrative performance has become of keen interest in recent years. While faculty evaluation instruments and policy systems are reasonably well established, the availability of administrator performance appraisal instruments, with established validity and reliability, is still relatively sparse.

The purpose of this letter is to invite and request your participation in a national study designed to establish validity and reliability data for an Administrative Effectiveness Survey (AES) originally developed at the University of Tennessee. This particular study will center primarily on the performance appraisal of academic deans. We would like to accent the idea, however, that the purpose of this study is not to evaluate the performance of deans on your campus, but to have faculty in different colleges complete the AES and to use those data in establishing reliability data for the survey. The AES will be reviewed by two focus groups for purposes of validating items in the instrument, and a factor analysis will also be completed to explore further validity issues.

We are asking for your participation in one of those focus groups and are enclosing the *Administrative Effectiveness Survey* for your review, which should only take 15 to 20 minutes. Your feedback will help to establish the content validity of this instrument. Please comment on the format and organization of the instrument, the response scale, the item construction, and provide suggestions for improvement to this survey.

Please return the survey and your comments in the envelope provided by [date]. While your participation is voluntary, we urge you to take the time to participate in this valuable study. Please feel free to contact us if you have any questions. We appreciate your cooperation.

Sincerely,

E. Grady Bogue, Ed. D.
Professor
865-974-6140
bogue@utk.edu

Kristi J. Nelms
Ed. D. Candidate
865-974-4578
knelms@utk.edu

APPENDIX D

LETTER TO INITIAL REVIEW PANEL

[Letterhead]

[Date]

[Name]

[Address]

Dear [Name]:

With the contemporary accent on accountability in American higher education, appraisal of both faculty and administrative performance has become of keen interest in recent years. While faculty evaluation instruments and policy systems are reasonably well established, the availability of administrator performance appraisal instruments, with established validity and reliability, is still relatively sparse.

The purpose of this letter is to invite and request your participation in a national study designed to establish validity and reliability data for an Administrative Effectiveness Survey (AES) originally developed at the University of Tennessee. This particular study will center primarily on the performance appraisal of academic deans. We would like to accent the idea, however, that the purpose of this study is not to evaluate the performance of deans on your campus, but to have faculty in different colleges complete the AES and to use those data in establishing reliability data for the survey. The AES will be reviewed by two focus groups for purposes of validating items in the instrument, and a factor analysis will also be completed to explore further validity issues.

We asking for you participation in one of those focus groups and are enclosing two versions of the *Administrative Effectiveness Survey* for your review, which should only take five to ten minutes. Each form has a different stem option for the response scale. In order to determine which response scale would be most appropriate for this instrument, please comment on the following:

- 1) Do you have a preference as to which response scale you would like to see on the questionnaire? Which response scale would you recommend?
- 2) Given the two versions of the questionnaire, would you rate the performance of your academic dean differently; that is, would you check a different response option for the same item on each questionnaire?

Please return the surveys and your comments in the envelope provided by [date]. While your participation is voluntary, we urge you to take the time to participate in this valuable study. Please feel free to contact us if you have any questions. We appreciate your cooperation.

Sincerely,

E. Grady Bogue, Ed. D.
Professor
865-974-6140
bogue@utk.edu

Kristi J. Nelms
Ed. D. Candidate
865-974-4578
knelms@utk.edu

APPENDIX E

LETTER TO PARTICIPATING INSTITUTIONS

[Letterhead]

[Date]

[Name]

[Address]

Dear [Name]:

With the contemporary accent on accountability in American higher education, appraisal of both faculty and administrative performance has become of keen interest in recent years. While faculty evaluation instruments and policy systems are reasonably well established, the availability of administrator performance appraisal instruments, with established validity and reliability, is still relatively sparse.

The purpose of this letter is to invite and request your participation in a national study designed to establish validity and reliability data for an Administrative Effectiveness Survey (AES) originally developed at the University of Tennessee. This particular study will center primarily on the performance appraisal of academic deans. We would like to accent the idea, however, that the purpose of this study is not to evaluate the performance of deans on your campus, but to have faculty in different colleges complete the AES and to use those data in establishing reliability data for the survey. The AES has already been reviewed by two focus groups for purposes of validating items in the instrument, and a factor analysis will also be completed to explore further validity issues.

If your institution would agree to participate, we would like to make the survey available to faculty in the following colleges: Agriculture, Arts and Humanities, Business, Education, Engineering, Law, Nursing, Social Work. Responses will be confidential and no individual respondent or dean will be identified in the study. Hard copies of the AES questionnaire could be sent direct to your office for distribution or direct to faculty members. However, the preferred method of participation desired would be to make the survey made directly available to faculty members via e-mail so that they can quickly respond on line. The estimated time for completion of the instrument is 15 minutes.

Participation via e-mail would require your authorization and e-mail addresses of your faculty. If you have a directory with e-mail addresses available, please send a copy of it or send an on line transmission of those e-mail addresses to knelms@utk.edu.

We very much hope that you would consent for [institution name] to participate. Please feel free to call or contact us if you have any questions. If you agree to participate, please complete the attachment and return it in the envelope provided by [date].

Sincerely,

E. Grady Bogue, Ed. D.
Professor
865-974-6141
bogue@utk.edu

Kristi J. Nelms
Ed. D. Candidate
865-974-4578
knelms@utk.edu

APPENDIX F

INSTITUTIONAL CONSENT TO PARTICIPATE FORM

INSTITUTIONAL CONSENT TO PARTICIPATE

The purpose of this study is to revise and to establish the reliability and validity of the *Administrative Effectiveness Survey* and to identify factors related to the performance of academic deans. The survey will take approximately 15 minutes to complete. Responses will be kept confidential. Results will be reported in aggregate form.

If you agree to participate, you will need to provide a list of the faculty members in the colleges listed below and their e-mail addresses so that they can be sent the website, a username, and password to access the evaluation instrument.

Do you agree to participate in this study?

____ Yes

____ No

How many faculty members are in each of the following colleges?

____ Agriculture

____ Arts and Humanities

____ Business

____ Education

____ Engineering

____ Law

____ Nursing

____ Social Work

YOUR NAME: _____ DATE: _____

SIGNATURE: _____

E-MAIL ADDRESS: _____

INSTITUTION NAME: _____

Please return this form by [DATE] (or as soon as possible).

We will need a list of your faculty members and their e-mail addresses. You can e-mail those to me at knelms@utk.edu, fax them to 865-974-6683 or mail them to me at 1021 Francis St. #131 Knoxville, TN 37916 (please let me know if you need a self-addressed envelope and I will send another one).

APPENDIX G

E-MAIL TO FACULTY MEMBERS

Dr. []:

Provost [] agreed for your institution to participate in a national study that I am conducting at the University of Tennessee, Knoxville. I have included a forwarded message from [] authorizing this study. I hope you will take a few minutes to assist with this research. The purpose of the study is to have faculty members at different colleges and universities complete a survey instrument of the performance of their academic dean so that the data may be used to establish the reliability of the instrument.

The Administrative Effectiveness Survey (AES) instrument is located on-line at:
[insert website]

I am providing you with a unique username and password to access the site. Please write them down and use them after you click on the above link (or you may have to copy and paste the link into your web browser). Once you access the site, you will find an informed consent letter to read. Upon reading and agreeing to participate in the study, you will be directed to the survey instrument. The on-line survey is designed to be completed in 15 minutes or less.

Username: []

Password: []

I would greatly appreciate your cooperation in this endeavor.

Sincerely,
Kristi Nelms
Ed.D. Candidate
865-974-4578
knelms@utk.edu

[forwarded message]

APPENDIX H

COVER LETTER TO FACULTY MEMBERS

(Informed Consent Letter)

[Letterhead]

[Date]

[Name]

[Address]

Dear [Name]:

With the contemporary accent on accountability in American higher education, appraisal of both faculty and administrative performance has become of keen interest in recent years. While faculty evaluation instruments and policy systems are reasonably well established, the availability of administrator performance appraisal instruments, with established validity and reliability, is still relatively sparse.

The purpose of this letter is to invite and request your participation in a national study designed to establish validity and reliability data for an Administrative Effectiveness Survey (AES) originally developed at the University of Tennessee. This particular study will center primarily on the performance appraisal of academic deans. We would like to accent the idea, however, that the purpose of this study is not to evaluate the performance of deans on your campus, but to have faculty in different colleges complete the AES and to use those data in establishing reliability data for the survey. The AES has already been reviewed by two focus groups for purposes of validating items in the instrument, and a factor analysis will also be completed to explore further validity issues.

As a faculty member in your college, you are in a position to observe and assess the performance of your academic dean. Your participation in this study is voluntary and you may decline to participate without penalty and you may withdraw from the study at anytime. If you decide to participate, you may withdraw from the study at anytime without penalty. If you withdraw from the study before data collection is completed, your data will be returned to you or destroyed.

If you agree to participate in the study, we would appreciate your cooperation in completing the survey, and in providing your honest feedback regarding the performance of your academic dean by [date]. The on-line survey is designed to be completed in fifteen minutes or less. This research should pose no more risk than you would encounter in everyday living. If you have any questions at any time about the study or the procedures, you may contact the researcher, Kristi Nelms, at 1021 Francis St. Knoxville, TN 37916, and 865-974-4201. If you have questions about your rights as a participant, contact the Research Compliance Services section of the Office of Research at the University of Tennessee at 865-974-3466.

The study is to be completed by [date] and will be available upon request from the authors. Responses will be confidential and no individual respondent will be identified in the study. Data will be stored securely and will only be available to the researchers of this study unless you specifically give permission in writing to do otherwise. No reference will be made in oral or written reports that could link you to the study.

The on-line survey can be found at the following website [url address] and your username is [] and your unique password to access the site is []. Please print your name and initial in the space provided that you have read the above information and agree to participate in this study. You may print a copy of this form for your records.

Please feel free to contact us if you have any questions. We appreciate your cooperation.

Sincerely,

E. Grady Bogue, Ed. D.
Professor
865-974-6141
bogue@utk.edu

Kristi J. Nelms
Ed. D. Candidate
865-974-4578
knelms@utk.edu

Consent: Participant name: _____ Initials: _____

APPENDIX I

FOLLOW-UP E-MAIL TO FACULTY

Dear Professor []:

A little over a week ago, I sent an e-mail asking for your help and this is a follow-up e-mail to again ask for your assistance. Dr. [], Provost and Vice President for Academic Affairs, agreed for your institution to participate in a national study that I am conducting at the University of Tennessee, Knoxville. I hope that you will take a few minutes to assist with this research. The purpose of the study is to have faculty members at different colleges complete a survey instrument of the performance of their academic deans so that the data may be used to establish the reliability of the instrument.

The Administrative Effectiveness Survey (AES) instrument is located on-line at: [website address].

I am providing you with a unique username and password to access the site. Please write them down and use them after you click on the above link. Once you access the site, you will find an informed consent letter to read. Upon reading and agreeing to participate in the study, you will be directed to the survey instrument. The on-line survey is designed to be completed in 15 minutes or less.

Username: []

Password: []

I would greatly appreciate your cooperation in this endeavor.

Sincerely,

Kristi Nelms
Ed.D. Candidate
865-974-4578
knelms@utk.edu

APPENDIX J
FINAL VERSION OF THE
ADMINISTRATIVE EFFECTIVENESS SURVEY

ADMINISTRATIVE EFFECTIVENESS SURVEY

Please assess the performance of the Dean of your academic college on the following effectiveness criteria.

Seriously Inadequate – Does not meet minimum expectations and termination should be considered

Needs improvement – Meets some minimum expectations and termination should not be considered

Satisfactory – Meets minimum expectations

Exemplary – Commendable performance, serves as an example to others

	Seriously Inadequate	Needs Improvement	Satisfactory	Exemplary	No Basis for Evaluation
PLANNING AND GOAL SETTING					
1. Establishes appropriate goals for self.	☐	☐	☐	☐	☐
2. Assists others in establishing appropriate goals.	☐	☐	☐	☐	☐
3. Establishes appropriate goals for the college.	☐	☐	☐	☐	☐
4. Encourages creation of a shared educational vision.	☐	☐	☐	☐	☐
5. Monitors progress towards plans and goals.	☐	☐	☐	☐	☐
6. Respects the university's mission.	☐	☐	☐	☐	☐
LEADERSHIP STYLE AND SKILL					
7. Adapts decision style to situation, as needed.	☐	☐	☐	☐	☐
8. Empowers others to take responsibility for decisions.	☐	☐	☐	☐	☐
9. Communicates effectively.	☐	☐	☐	☐	☐
10. Balances consensus and conviction in decision-making.	☐	☐	☐	☐	☐
11. Responds effectively to complaints.	☐	☐	☐	☐	☐
12. Manages conflict effectively.	☐	☐	☐	☐	☐
13. Represents the college with credibility.	☐	☐	☐	☐	☐
14. Motivates colleagues to high performance standards.	☐	☐	☐	☐	☐
RESOURCE STEWARDSHIP					
15. Exhibits good judgment in personnel selection.	☐	☐	☐	☐	☐
16. Exhibits good judgment in personnel retention and improvement.	☐	☐	☐	☐	☐
17. Effectively manages resources entrusted to the care of the office/university.	☐	☐	☐	☐	☐
18. Encourages others to be effective resource stewards.	☐	☐	☐	☐	☐
19. Effectively prioritizes use of resources.	☐	☐	☐	☐	☐
VALUES AND STANDARDS					
20. Defines and expects high standards of performance in programs.	☐	☐	☐	☐	☐
21. Treats colleagues with respect.	☐	☐	☐	☐	☐
22. Exhibits an appropriate balance of compassion and courage in relationships.	☐	☐	☐	☐	☐
23. Reflects personal integrity.	☐	☐	☐	☐	☐
24. Emphasizes and supports diversity in university life.	☐	☐	☐	☐	☐

	Seriously Inadequate	Needs Improvement	Satisfactory	Exemplary	No Basis for Evaluation
25. Candidly shares information, as appropriate.	☐	☐	☐	☐	☐
26. Builds a climate of trust and credibility.	☐	☐	☐	☐	☐
27. Challenges ineffective practices.	☐	☐	☐	☐	☐
ACADEMIC STEWARDSHIP					
28. Fosters excellence in teaching methods and research.	☐	☐	☐	☐	☐
29. Effectively represents faculty issues to university administration.	☐	☐	☐	☐	☐
30. Effectively builds relationships with external stakeholders.	☐	☐	☐	☐	☐
31. Serves as an effective external raiser of funds for the college.	☐	☐	☐	☐	☐
32. Effectively models scholarship by publishing and/or presenting papers regularly.	☐	☐	☐	☐	☐
33. Encourages faculty involvement in community projects and service endeavors.	☐	☐	☐	☐	☐
34. Emphasizes scholarly development and research among faculty members.	☐	☐	☐	☐	☐
OVERALL EVALUATION:	☐	☐	☐	☐	☐

If you were to use this instrument on your campus, would you include the following open-ended questions?

- ☐ Yes 1) Describe briefly one action, behavior, or situation in which the Dean exhibited commendable and exemplary leadership.
- ☐ No
- ☐ Yes 2) Describe briefly one action, behavior, or situation in which the Dean's performance reflected a need for improvement.
- ☐ No

What other questions would you include in this instrument?

What recommendations would you make for improving this evaluation instrument?

Please check the appropriate box:

Gender: ☐ Female
☐ Male

Primary Race: ☐ African American
☐ Asian
☐ Hispanic
☐ Caucasian
☐ Pacific Islander
☐ Other

College: ☐ Agriculture
☐ Arts and Humanities
☐ Business
☐ Education
☐ Engineering
☐ Law
☐ Nursing
☐ Social Work
☐ Other

APPENDIX K

REVISED *ADMINISTRATIVE EFFECTIVENESS SURVEY*
WITH THREE FACTORS

ADMINISTRATIVE EFFECTIVENESS SURVEY

Please assess the performance of the Dean of your academic college on the following effectiveness criteria.

Seriously Inadequate – Does not meet minimum expectations and termination should be considered

Needs improvement – Meets some minimum expectations and termination should not be considered

Satisfactory – Meets minimum expectations

Exemplary – Commendable performance, serves as an example to others

	Seriously Inadequate	Needs Improvement	Satisfactory	Exemplary	No Basis for Evaluation
LEADERSHIP					
1. Establishes appropriate goals for self.	☐	☐	☐	☐	☐
2. Encourages creation of a shared educational vision.	☐	☐	☐	☐	☐
3. Monitors progress towards plans and goals.	☐	☐	☐	☐	☐
4. Adapts decision style to situation, as needed.	☐	☐	☐	☐	☐
5. Empowers others to take responsibility for decisions.	☐	☐	☐	☐	☐
6. Communicates effectively.	☐	☐	☐	☐	☐
7. Balances consensus and conviction in decision-making.	☐	☐	☐	☐	☐
8. Responds effectively to complaints.	☐	☐	☐	☐	☐
9. Manages conflict effectively.	☐	☐	☐	☐	☐
10. Exhibits good judgment in personnel selection.	☐	☐	☐	☐	☐
11. Exhibits good judgment in personnel retention and improvement.	☐	☐	☐	☐	☐
12. Treats colleagues with respect.	☐	☐	☐	☐	☐
13. Exhibits an appropriate balance of compassion and courage in relationships.	☐	☐	☐	☐	☐
14. Reflects personal integrity.	☐	☐	☐	☐	☐
15. Candidly shares information, as appropriate.	☐	☐	☐	☐	☐
16. Builds a climate of trust and credibility.	☐	☐	☐	☐	☐
17. Challenges ineffective practices.	☐	☐	☐	☐	☐
EXTERNAL ACADEMICS					
18. Effectively builds relationships with external stakeholders.	☐	☐	☐	☐	☐
19. Serves as an effective external raiser of funds for the college.	☐	☐	☐	☐	☐
20. Effectively models scholarship by publishing and/or presenting papers regularly.	☐	☐	☐	☐	☐
21. Encourages faculty involvement in community projects and service endeavors.	☐	☐	☐	☐	☐
22. Emphasizes scholarly development and research among faculty members.	☐	☐	☐	☐	☐
MISSION, STANDARDS, AND RESOURCES					
23. Respects the university's mission.	☐	☐	☐	☐	☐
24. Motivates colleagues to high performance standards.	☐	☐	☐	☐	☐
25. Effectively manages resources entrusted to the care of the office/university.	☐	☐	☐	☐	☐
26. Encourages others to be effective resource stewards.	☐	☐	☐	☐	☐

	Seriously Inadequate	Needs Improvement	Satisfactory	Exemplary	No Basis for Evaluation
27. Effectively prioritizes use of resources.	☐	☐	☐	☐	☐
28. Defines and expects high standards of performance in programs.	☐	☐	☐	☐	☐
OVERALL EVALUATION:	☐	☐	☐	☐	☐

If you were to use this instrument on your campus, would you include the following open-ended questions?

- ☐ Yes 1) Describe briefly one action, behavior, or situation in which the Dean exhibited commendable and exemplary leadership.
- ☐ No
- ☐ Yes 2) Describe briefly one action, behavior, or situation in which the Dean's performance reflected a need for improvement.
- ☐ No

What other questions would you include in this instrument?

What recommendations would you make for improving this evaluation instrument?

Please check the appropriate box:

Gender: ☐ Female
☐ Male

Primary Race: ☐ African American
☐ Asian
☐ Hispanic
☐ Caucasian
☐ Pacific Islander
☐ Other

College: ☐ Agriculture
☐ Arts and Humanities
☐ Business
☐ Education
☐ Engineering
☐ Law
☐ Nursing
☐ Social Work
☐ Other

APPENDIX L

**SUMMARY OF DEMOGRAPHIC INFORMATION
FOR THE NATIONAL STUDY**

Gender by Race

	Race	Caucasian*	African** American	Asian	Hispanic	Other	Not*** Identified	Total
Gender								
Male		86	1	5	1	6	1	100
Female		66	1	0	2	2	3	74
Total		152	2	5	3	8	4	174

*2 respondents identified as Caucasian, but with no associated gender

**1 respondent identified as African American, but with no associated gender

***2 respondents did not identify and did not specify gender

College Type by Gender*

	Gender	Male	Female	Total
College Type				
Agriculture		3	2	5
Arts & Humanities**		35	16	51
Arts & Sciences		10	4	14
Business		16	6	22
Education		15	15	30
Engineering		8	1	9
Human Ecology		0	1	1
Humanities		0	1	1
LEC		1	0	1
Library		0	1	1
Music		1	0	1
Nursing		0	19	19
Other		6	7	13
Sciences		1	0	1
Special Programs		1	0	1
Theology		3	0	3
Total		100	73	173

*2 respondents did not identify college type or gender

**3 respondents identified as Arts and Humanities, but with no associated gender

Gender by Carnegie Type

Gender	Carnegie Type	Master's I*	Master's II**	Unidentified***	Total
Male		35	28	37	100
Female		33	15	26	74
		68	43	63	174

*2 identified as Master's I, but with no associated gender

**1 identified as Master's II, but with associated gender

***2 did not identify Carnegie type or gender

APPENDIX M

**SUGGESTIONS FOR ADDITIONAL QUESTIONS TO BE ADDED
TO THE INSTRUMENT**

- Do the systems in place at your institution allow for enough communication between Dean and faculty? Do you feel comfortable meeting with your Dean without the knowledge of your Chair? Have you found any opportunities to develop a professional relationship with your Dean? Do you feel your Dean has any knowledge of what you do?
- Possibly a question about whether the Dean treats colleagues with fair judgment (under leadership style) or too much bias toward "friends" on the faculty.
- Knowledge of academic area being supervised.
- Questions are adequate.
- Open-ended questions on strengths and weaknesses.
- How long have you been employed at this institution? (I have only been there for a year, so I had difficulty answering many questions and will probably have different responses than colleagues who have been there longer.)
- I would like to see a question or questions dealing with fairness of action and judgment in dealing with faculty and staff. This might be included in a section on interpersonal relationships with faculty and staff.
- Space for additional comments by the evaluator.
- I think this fits under handles complaints, but maybe something about interaction with students and ensuring they follow chain of command.
- Maintains confidentiality of faculty.
Supports faculty actions related to curricular issues.
Supports faculty decision making relative to academic decision making.
- Uses skills of administrative team to build cohesive workforce.
Represents priorities appropriately in budget.
- Questions related to conflict resolution among faculty--how to resolve.
Question related to setting organizational behavioral standards within a College so that non-cooperative faculty know that there are boundaries or standards within the process.
- Under Academic Stewardship: Encourages interdisciplinary/inter-institutional collaboration for scholarly endeavors.

- How many times during the academic year do you have occasion to meet with this dean?
- Distributes financial resources equitably between disciplines.
Exemplifies a professional attitude.
According to your personal expectation, represents the institution in a manner befitting a university administrator.
- Perhaps questions about specific issues that concern many if not most colleges, such as contingent/adjunct labor, salary equity and fairness among faculty in the college or across the university, demonstration or utilization of current knowledge of administrative scholarship.
- There are too few questions that deal with development of academic scholarship among faculty members. There needs to be questions about how appropriate the standards at the level of Dean are for tenure and promotion.
- Ask for suggestions for improvement, or for reporting certain situations.
- Modify question 1 above to read: "Describe briefly one action, behavior or situation in which the dean exhibited commendable and exemplary leadership in assisting the faculty in meeting its goals within the university."
- There could possibly be some reference in questions as to the freedom in decision making as dictated by finances, facilities, or top administration which often control, to some extent, the actions of the dean regardless of their wishes. eg. hiring personnel, providing support, etc.
- Does the Dean demonstrate a knowledge of the operations and functioning of each of the academic units within his/her college or school?
- Under Leadership Style and Skill: Promotes shared governance.

APPENDIX N
RECOMMENDATIONS TO IMPROVE THE INSTRUMENT

- Perhaps add "above average" between satisfactory and exemplary.
- Allow room for specific and personal concerns to be addressed.
- Leave as is....
- As an Assistant Professor with limited experience, it is difficult to answer many of these questions.
- Allow comments for responses to Likert-type scale questions.
- I would request a comment for any score of "Seriously Inadequate" or "Exemplary" to give the Dean more specific information concerning the item in question.
- Provide space for comments and reflection.
- None.
- I think it is a great instrument - it gets at the issues; the questions I stated above may already be subsumed in other questions
- I am not certain that the word exemplary fits most situations that might be above average. That category would be difficult to select (in my opinion).
- I think the above statement would give more reliable data since some people who are asked to evaluate deans could have more exposure and some a lot less, or you might target those people who work more closely with a particular dean. I have been asked to evaluate people with whom I have had very little direct contact and I find this difficult and feel that the information I give provides inaccurate overall data.
- I like brief but not sure this is a valid instrument for faculty to use in rating their dean. Perhaps those in administration would serve better as the evaluator.
- The format is great and should really assist you with a respectable response rate. It is quick and fast and this should also assist with obtaining a 60% response rate. I would assume that you are doing a multiple regression analysis on your dependent variable. If so, this looks like a quite valid instrument. Good job!
- I would make the two open-ended questions optional.

- A larger proportion of the questions should deal with academic development, toward both teaching and research. Especially, however, more questions to address research expectations from faculty.
- The gap between "satisfactory" and "exemplary" seems too extreme. Should there not be a category between these two?
- Need questions that call for narrative answers, not just ratings.
- Modify question 2 above to read: "Describe briefly one action, behavior or situation in which the dean's performance reflected a need for improvement in preventing the faculty from meeting its goals within the university."
- None.
- I don't have enough direct, firsthand knowledge to separate decisions made by the Dean and those made by others in the Office of the Dean, like the Associate Dean. So, separating the Dean from the Office of the Dean and its staff was difficult on some questions. I'm not sure what changes (if any) that concerns implies.
- Question #1 "personal goals" might be difficult to validate or measure with any accuracy even with personal friends.
Question #31 fund raising, depending on the institution should normally be handled by other individuals or departments, presidents, development etc.
It may even be restricted or limited apart from the above question
- You are assuming that the Dean has discretion in making decisions. I have seen what the Dean has to deal with in Presidents, Vice-Presidents, Associate Vice-Presidents, Assistant Vice-Presidents, etc. You are constantly protecting your school from idiots with power but no responsibility. I find it difficult to evaluate the Dean when he is forced to support incorrect decisions from others.
- Neutral midpoint in scale.
- The choices in the "Your Carnegie classification" probably need further explanation. We do have doctoral programs with three different majors; however, I personally do not view us as doctoral extensive or intensive.
- This instrument was very well done and was easy-to-read and understand.
- None.

VITA

Kristi Jolena Nelms was born December 3, 1968 in Norman, Oklahoma. She attended elementary schools in Bridge Creek and Chandler, Oklahoma and graduated from Chandler High School in 1987. She went on to attend Oklahoma State University in Stillwater, Oklahoma and earned a Bachelor of Science in Mathematics, with a minor in Spanish and a secondary education teaching certificate in 1992. She then moved to Knoxville, Tennessee and began a master's program in the fall of 1992. She graduated with a Master of Science in College Student Personnel from the University of Tennessee, Knoxville in 1994 and then, in August 2005, received the degree of Doctor of Education in Educational Administration and Policy Studies, with a minor in Statistics.

Kristi has worked in University Housing at the University of Tennessee since 1992. She is currently an Assistant Director in the department.

4535 0545 28
11/09/05 MAB