

Evaluating Department Chair and Student Leadership in Higher Education

Linda P. Thurston
and B. Jan Middendorf

Introduction

In higher education, assessment of leadership capacities and performance of department chairs and students allows stakeholders to evaluate individuals and programs. To that end, this article describes the Individual Development and Educational Assessment (IDEA) Feedback for Department Chairs system, a psychometrically sound instrument developed for evaluating department chairs at Kansas State University, and the process used to develop a contextually based leadership assessment instrument for students in the university's Leadership Studies program. After collecting data regarding leadership of chairs or students, the evaluation process uses a standard or benchmark placing value or merit on the factors measured. The article begins with a background section that presents a framework for accountability in higher education leadership followed by a subsection that defines and compares the concepts of assessment and evaluation. The third section contains a review of relevant literature

Linda P. Thurston is Program Director for the National Science Foundation in Washington, D.C. Her responsibilities include evaluation and work to reach under-represented groups in science, technology, engineering, and mathematics, specifically with regard to disability and gender. She is also the founding director of the Office of Educational Innovation and Evaluation at Kansas State University, and was Assistant Dean of the College of Education and Professor of special education at Kansas State University. She has been nationally recognized for her work with poverty level single mothers and youth. In 2002, she received the Michael C. Holen Outstanding Leadership Award presented for embodying the School of Leadership Studies mission to develop knowledgeable, ethical, caring, inclusive leaders for a diverse and changing world.

B. Jan Middendorf is Acting Director of the Office of Educational Innovation and Evaluation at Kansas State University and Research Assistant Professor. Her research interests include leadership, evaluation, and curriculum instruction. As Acting Director, she provides leadership in project development and the creation of collaborative relationships with faculty and funding agencies for the purpose of developing innovative joint projects. She has extensive expertise in project design, implementation, management of program evaluations, and strategic planning for projects.

on establishing indicators for evaluating leadership in higher education, context for assessing leadership, and theoretical base. In the fourth section, assessment instruments and evaluation methods are described. The article closes with a summary section.

Background

A Framework for Accountability in Higher Education Leadership

In *Pursuing Excellence in Higher Education*, Ruben (2004) listed eight critical challenges to higher education based upon the Malcolm Baldrige National Quality Award Program framework for organizational excellence. Of these, three are relevant to assessment and evaluation of leadership in higher education:

- Integrating assessment, planning, and improvement;
- Becoming a more effective learning organization;
- Devoting more attention and resources to developing exceptional educational leadership leaders.

With regard to evaluation in higher education, an element of a thriving learning organization, Ruben (2004) listed six major functions:

1. Accountability. Programs are accountable to funders and/or administrators. Evaluation provides answers to these questions: Is the program or organization doing what it says it is doing? Are the activities and outcomes of the organization congruent with its mission? Are students learning what faculty are expecting them to learn?
2. Program/continuous improvement. Evaluation data provide feedback to programs that informs modifications to better serve stakeholders or meet goals. Accrediting bodies want to know that programs are continuously improving their operations and outcomes.
3. Dissemination/replication. Evaluation can address the following important question: Is a program ready to be disseminated to others? For example, is a faculty development program in leadership worth replicating in other years or in other colleges?
4. External funding/continued support. Can program organizers demonstrate why it is worthy of receiving external support from funders? For example, can a leadership development program demonstrate that its funded program is being conducted as proposed and that it is making progress toward developing skilled and ethical student leaders?
5. Rationale for ongoing stakeholder support. Stakeholders want to know that their needs are being met and that their time, expertise, and funds are being used to produce the outcomes they expect. For example, did a leadership institute produce enough expected changes in participants to warrant continued support by university administration?
6. Capacity building within higher education institutions for assessment and reflection. Evaluation forces units and programs to begin developing their own resources to include ongoing evaluation. This contributes to a culture of accountability and the internal capacity to assess and evaluate programs and products, leading to a more effective learning organization.

These six functions relate directly to issues of leadership in higher education and provide a framework for accountability. Department chairs and students are both subjects of evaluation (the evaluands) and consumers of evaluation results.

Assessment vs. Evaluation

Assessment is the process of defining variables to be measured; designing or selecting the metrics for gathering the information about those variables; and collecting credible data using appropriate methodology. *Evaluation* is the process of determining the value, merit, or worth of a program or personnel.

Assessment of outcomes does not by itself produce enough evidence to permit a thorough understanding of programs, policies, and individuals in higher education. Evaluation uses information based on credible evidence generated through assessment to make judgments of relative value. Assessment indicates what results have been produced, but it does not determine causation, indicate how those results were achieved, or compare those results with accepted higher education standards. Therefore, evaluators utilize accepted evaluation designs or established standards for the process of establishing the value of merit of the evaluand.

Evaluation is a vibrant and engaging activity that leads to powerful learning and well-informed action (Hannum, Martineau, & Reinelt, 2007). Evaluation has two arms: accumulating and summarizing data; and drawing conclusions about the value or relevance of standards in a program (Scriven, 1991). The specific form and scope of an evaluation depend on its purposes and audience, the nature of the evaluand, and the organizational context within which the program/individual operates. However, higher education presents a unique context in which to conduct assessment and evaluation. Contextual issues in evaluating leadership in higher education are discussed in a later section.

Evaluation facilitates decision-making when it combines sound procedures with issues valued by stakeholders. The selection of variables to measure, the measurement tools, and the evaluation design depends on the types of decisions to be made. Therefore, an evaluator begins with questions, such as: What is the purpose of the evaluation? What is the mission of the institution? What are the program or project goals? What are the expected outcomes? What are the criteria for success? What is the role of the individual in the institution, and what are the expected competencies attributed to that role? What decisions need to be made?

Approaching issues from an evaluative perspective enables one to consider multiple perspectives and draw lessons as a natural part of the way work is done (Hannum, Martineau, & Reinelt, 2007). This perspective contributes to developing and sustaining an effective learning organization (Ruben, 2004). Evaluation equals assessment plus a judgment related to the value of a program, employee, or process. Evaluation of leadership in higher education, therefore, includes two essential elements: assessment of leadership; and establishment of a standard with which to compare the results of the assessment.

Review of Literature

Establishing Indicators for Evaluating Leadership in Higher Education

One of the biggest challenges in evaluation is choosing what kind of information best answers the questions posed. It is important to have general agreement across target audiences on what success looks like. Indicators are the starting point for data collection and reporting, and are selected to represent important outcomes or performance measures. Therefore, consideration of indicators is an essential element of evaluation in higher education.

Much has been written about the indicators of successful leadership (Stufflebeam, 1999). Because individuals are the focus in evaluating leadership, leadership indicators, for the most part, relate to the traits, skills, behaviors, attitudes, values, competencies, and knowledge. Also, specific contextual variables such as collaboration, cultural competence, relationship building, problem solving, empowerment of others, catalyzing, and sustaining change are possible indicators that could be evaluated in leaders or potential leaders in higher education settings. These potential indicators are contextually bound in higher education. For example, a department chair might be evaluated on her or his ability to empower faculty in the department. However, this might not be an indicator of success for a student in a leadership studies program. For example, a more likely indicator for a student would be knowledge of leadership theories.

EvaluleAD methodology for evaluating leadership development activities identifies fundamental parameters that include context, domains, and result types (indicators) (Grove, Kiber, & Haas, 2005). Wisniewski (1999) examined leadership competence models to find a model that fit with higher education in general and extension services specifically. The four models had significant overlap in their categorizations of leadership competencies; however, the discrepancies led Wisniewski to generate a leadership competence model specifically for the university extension context. She used grounded theory methodology and the critical incident technique in her research. Her results were seven leadership categories and related abilities: (1) core set of values and vision; (2) effective communication; (3) reflection and analysis; (4) positive climate; (5) facilitation and collaboration; (6) problem solving and risk taking; and (7) perseverance. These included a short list of indicators for each category. For example, positive climate, included the ability to interact comfortably with a variety of people, establish a high-trust environment, develop a sense of empathy, and motivate and inspire others. Wisniewski utilized these indicators as the basis for a leadership education program for leaders at their university system.

Defining indicators for measuring leadership in collegiate students has been ongoing work for the W.F. Kellogg Foundation and others. The Council for Academic Standards in Higher Education (CAS) established 16 Student Learning & Development Outcome Domains for student leadership development programs (Miller, 2003). (See the textbox on the next page for a listing of these). As with Wisniewski's (1999) categories of leadership indicators, each of the CAS standards includes a list of examples of achievement indicators for each category. For example, indicators related to "clarified values" are: articulates personal values; acts in congruence with personal values; makes decisions that reflect personal values; demonstrates willingness to scrutinize personal beliefs and values; and identifies personal, work, and lifestyle values and explains how they influence decision-making.

Context for Assessing Leadership

Context is an important consideration in establishing indicators of successful leadership. The concept of context recognizes that leadership may assume a wide variety of forms and expressions of personal and cultural style. Contextual factors include opportunities, management systems, expectations of others, and institutional culture (Peters & Baum, 2007). In their work with the Sustainable Leadership Initiative funded by W.K. Kellogg and USAID Grove, Kiber, and Hass (2005) outlined two fundamental steps in evaluation planning: defining the context of leadership to be evaluated; and defining the

**Student Learning and Development Outcome
Domains for Student Leadership
Development Programs**

- **Intellectual Growth**
- **Effective Communication**
- **Enhanced Self-Esteem**
- **Realistic Self-Appraisal**
- **Clarified Values**
- **Career Choices**
- **Leadership Development**
- **Healthy Behavior**
- **Meaningful Interpersonal Relationships**
- **Independence**
- **Collaboration**
- **Social Responsibility**
- **Satisfying and Productive Lifestyles**
- **Appreciating Diversity**
- **Spiritual Awareness**
- **Personal and Educational Goals**

Source: Miller (2003).

domains of impact. The Wisniewski (1999) study and the CAS standards both showed responsiveness to the context of higher education in their domains or categories.

Mitchell (2004) also emphasized the importance of the consideration of context in her discussion of assessment and evaluation of department chairs. Although some indicators for success of department leaders were common to all faculty, e.g. service to the college and university, other aspects of the business of running a department, such as evaluation of faculty and growth of an academic department, were viewed as unique. Usually written guidelines in the form of university policy related to the context of the job and the specific domains included.

Theoretical Base

Montez (2003) utilized significant stakeholder input to develop a five-dimensional theory of higher educational leadership:

1. Integral
2. Relational
3. Credibility
4. Competence
5. Direction/guidance

This led to the development of the Higher Education Leadership Inventory (HELI) to assess the attributes or behaviors considered to be necessary for effective leadership in higher education (Montez, 2003).

Assessment Instruments and Evaluation Methods

Traditional leadership assessment instruments overlook the specific context of higher education, providing little systematic knowledge for higher education administrators about behaviors, leadership styles, and effectiveness in higher education (McDade, 1987; Williams, 2001). Unique aspects of the higher education environment include:

shared governance; autonomy and academic freedom of faculty; synergism of expectations for research; teaching; and service; and leadership. There have been few research studies related to appropriate behaviors and attributes of persons for leading in this unique environment (Montez, 2003). Choosing methods or developing instruments to assess leadership depends on the kind of leadership indicators to be measured. For example, if an important indicator of successful leadership for a university administrator is communicating a vision for the unit, then a potential measurement method could be an interview during which the administrator is asked to describe her vision.

After domains and indicators for leadership have been established, good measures have to be adopted or developed. Unbiased instruments or methods that are appropriate measures of performance and produce a reasonable level of objective reliability are essential. Poister (2003) listed these criteria for useful performance measures:

- Valid and reliable
- Meaningful and understandable
- Balanced and comprehensive
- Clear regarding preferred direction of movement
- Timely and actionable
- Resistant to goal displacement
- Cost-sensitive

Instruments that do not attend to these criteria produce unreliable and invalid data. A favorite expression of evaluators related to poor instrument design is “garbage in, garbage out.” For example, survey items that are unclear or that incorporate biases can lead to serious measurement problems. Vague, double-barreled,¹ or ambiguous interview questions lead to problems because respondents are likely to interpret them in different ways. Leading questions in a focus group can unintentionally prompt respondents to answer in a certain way.

The choice of assessment methods should be determined by what indicators are chosen. The use of multiple methods is common in evaluating leadership in higher education. These include surveys, interviews, journals, observation, focus groups, and tracking accomplishments, e.g. publications, presentations, and community leadership positions held. In addition, a 360-assessment is frequently used. Here colleagues and coworkers of a university chairperson are interviewed or surveyed. For student assessment, mentors, faculty, advisors, supervisors or peers might be included in the evaluation.

The two most common methods used to assess leadership in higher education are standardized commercial instruments or “home grown” instruments that are based on the context of the situation and the unique indicators for specific role expectations. Relying on instruments with established, well-researched psychometric characteristics assures the accurate and appropriate measurement of leadership in the settings for which the instruments were developed. When choosing such instruments, reliability and validity must be considered. Most instruments report their reliability; that is, the degree to which the instrument is consistent. Reliability estimates of .80 are considered acceptable (Kline, 1999). Validity refers to the fit of an instrument to a situation and answers the question: Does the instrument measure what it is expected to measure? Both reliability and validity are essential considerations in choosing an instrument to assess leadership to assure the veracity of data collected.

Because of the wide range of definitions, domains, and situations related to leadership in higher education, many evaluators choose to develop their own instruments. For example, Montez (2003)

examined five psychometrically sound leadership assessment instruments that measured leadership attributes, practices, and skills; used multi-rated instrument; and had been tested on higher education populations. However, she found that none fit the domains of leadership in higher education.

*Department Chairs in Higher Education:
Assessing and Evaluating Leadership*

The roles and responsibilities of academic department chairs have always been a challenge given the complexity of their role as negotiator, facilitator, evaluator, and administrator of faculty who have a great deal of autonomy. In addition, most department chairs enter into these positions with little awareness of what the job really entails and even less preparation for what awaits them in the position (Gmelch & Miskin, 1993, 2007; Wheeler, Seagren, Becker, Kinley, Mlinek, & Robson, 2008).

Research from Wheeler et al. (2008) indicates that the role of department chairs has become more critical as an agent of change. Moreover, the importance of department chair effectiveness in terms of leadership and accountability has become salient in recent years. The need to make departments stronger, more effective, and efficient through department chair leadership is increasing as is the need to understand how to assess these efforts (Leaming, 2007). With a focus on improving effectiveness and enhancing accountability, department chairs need a comprehensive evaluation process to assess how well they are performing in their positions.

The Individual Development and Educational Assessment (IDEA) Center at Kansas State University developed the IDEA Feedback for Department Chairs system for evaluating and developing department chairpersons. The original instrument, the Departmental Evaluation of Chairperson Activities for Development (DECAD), was first made available in 1977. In 1999, it was revised to reflect the literature on department chair leadership and effectiveness and given its current name. The system is comprised of two instruments and a summary feedback report: The Faculty Perceptions of Department Head/Chair Survey (FPDHS); and the Chair Information Form (CIF).² The system is designed to measure effectiveness for both summative evaluation, i.e., recommendations regarding merit salary, promotion, and other administrative decisions, and formative evaluation, i.e., improving administrative performance. This is accomplished by soliciting faculty input on how well the department chair has used different administrative methods to fulfill responsibilities he or she identifies as important or essential for the department. Results from the two instruments are analyzed and then summarized in the Feedback for Department Chair Report.³

The FPDHS is a 70-item instrument containing 67 objectively worded items and 3 short-answer written-response items.⁴ All objective items were constructed using a Likert-type format with five possible responses ranging from 1 to 5 (1=low; 5=high); however, the wording of the scale anchors varies depending on the subscales. In the first 20 items on the FPDHS instrument, faculty rate their respective department chair's performance on various administrative responsibilities. Five apriori subscales are assumed for administrative responsibilities: (1) administrative support; (2) personnel management; (3) program leadership/support; (4) building image/reputation; and (5) developing positive climate. The scale for these items ranges from 1 to 5 (1=poor; 5=outstanding).

For items 21-30, faculty rate the department chair's strengths and weaknesses on personal characteristics. Five apriori subscales are assumed for personal characteristics: (1) ability to resolve issues; (2) communication skills; (3) steadiness; (4) trustworthiness; and (5) openness. The scale for these items ranges from 1 to 5 (1=definite weakness; 5=definite strength). Faculty also indicate how frequently their department chair performed administrative behaviors associated with five apriori subscales: (1) democratic/humanistic; (2) goal-oriented/structured; (3) supports faculty; (4) promotes positive climate; and (5) promotes department advancement. These scales include subsets of items 31-60 where the scale ranges from 1 to 5 (1=hardly ever; 5=almost always).

Items 61-65 refer to financial, bureaucratic, and faculty impediments to the chair's effectiveness. The scale for these items ranges from 1 to 5 (1=definitely false; 5=definitely true). Items 66-67 use the same scale and are designed to provide a summary judgment of the department chair. Item 66 states, "I believe the department would be better off if we replaced the current department chair," and Item 67 states, "I have confidence in the department chair's ability to provide leadership to the department." Items 68-70 are open-ended questions related to suggestions for improvement and areas to strengthen from the faculty's perspective.

The CIF is comprised of 30 items including 20 questions that ask department chairs to rate various administrative responsibilities on importance, ranging from 1 to 5 (1=not important; 5=essential). The remaining 10 items query department chairs about various departmental characteristics. On the FPDHS, faculty rate their respective department chair's performance on each of the same 20 responsibilities described above (items 1-20).⁵

The resulting Feedback for Department Chair Report contains individualized data along with national comparisons that provide direction on specific areas of strength and strategies for improvement. The report provides both summative and formative feedback. The summative portion of the feedback report is designed to accommodate differences among departments by developing individualized "priority profiles." The priority profiles are based on the ratings from the faculty on the relative importance of responsibilities commonly stressed by academic departments. These standards are used to weight faculty ratings of how well each responsibility was performed. The weighted averages are used as the principal measure of administrative effectiveness, (Hoyt, Bailey, Pallett, & Gross, 1999). In order to provide assistance in improving performance, strengths and weaknesses are diagnosed by comparing ratings from the national database with regard to "relevant administrative behaviors" with the ratings from the faculty respondents from that specific department. The domains or indicators of interest are based on the apriori subscales within the instruments that reflect the essential behaviors, characteristics, and methods for effective administrators described in the majority of department chair literature.

Middendorf, Benton, and Webster (2009) examined the validity and reliability of the FPDHS and CIF. Overall, they found strong evidence for the reliability, construct validity, and concurrent validity of three underlying dimensions that department chairs deemed most important: foster faculty talents; develop collegiality; and improve the department's campus reputation. Other elements of importance included communicating department needs, guiding curriculum development, and orienting new faculty and staff. Based on this

research and several focus groups held with department chairs, the IDEA Center is in the process of revising the FPDHS system.

The FPDHS is the only nationally normed instrument for evaluating department chairs, and it provides a formative basis for their development. The survey takes into account that different management styles and strategies come into play when addressing different responsibilities. Measures of effectiveness are based on faculty input on how well the chair has used different administrative methods to meet identified goals for the department (Hoyt et. al, 1999). This mechanism allows the department the flexibility of analyzing results that are relevant to the department chair's performance and the faculty's perception of his or her performance. Because the standards are based on national norms and effective practice, they provide appropriate guidance for professional development and, ultimately, improved performance.⁶

Assessing and Evaluating Student Leadership at Kansas State University

Binard and Brungardt (1997) noted that little guidance exists related to assessment within undergraduate leadership programming and point out the need for assessment procedures to measure leadership growth in student development. An example of a standardized commercial instrument for student leadership assessment is the Leadership Practices Inventory, an instrument developed for a 360-degree assessment with a 5-point Likert-type survey based on a 5-factor framework (Kouzes & Posner, 1988). Kouzes and Posner developed the framework for their instrument based on interviews and case studies of over 1,000 corporate managers. The instrument shows internal reliability with an alpha coefficient between .70 and .85 (Posner & Kouzes, 1992). Although this instrument does not have the history of the set of instruments for department chairs, it was found to be helpful in assessing student leadership in several studies. In their study of 27 students at a community college, Binard and Brungardt (1997) utilized a pre-post evaluation design and the Leadership Practices Inventory. Brungardt and Crawford (1996) utilized the LPI-Self instrument as well as an attitude survey and a knowledge examination to assess students in a leadership development program.

As part of a comprehensive evaluation of the Leadership Studies Program at Kansas State University, a contextually appropriate student leadership assessment instrument was developed. Surveys typically ask participants to rate the effect of a program on a set of indicators. To establish student leadership indicators, an alumni survey team utilized input from many groups of stakeholders that included faculty, advisory board members, and others involved in the program.⁷ The evaluation team worked with stakeholders to determine areas in which student change can be expected and linked to the mission of the unit. Once domains and indicators were identified from this process, an appropriate and accurate measure for assessing student leadership was developed and implemented.

Multiple methods of data collection were used to examine the perceptions of students who progressed through the series of four courses required for the minor in Leadership Studies. Surveys were conducted for three of the four courses: Introduction to Leadership Concepts; Culture and Context in Leadership; and Leadership in Practice. For the final course, Senior Seminar in Leadership Studies, focus groups were conducted.

For Introduction to Leadership Concepts, survey questions related to student expectations for the course and the minor. This survey instrument consisted of scaled and open-ended items as well as demographic questions. Forty scaled items assessed the extent to which students believed they had achieved various leadership and learning outcomes. The open-ended items provided students with the opportunity to share expectations of outcomes or benefits from their experiences in the program. These responses were analyzed, and the results were combined with the results of the previous solicitations for information from stakeholders, program learning objectives, mission, and literature related to student leadership indicators. The result was a set of leadership skills and competencies that were grouped into four domains: critical thinking; knowledge about leadership theories and practices; communication and collaboration; and diversity. For the senior seminar, two questions framed the focus group discussion: (1) What are the benefits of participating in the Leadership Studies Program? and (2) What is the value of earning a minor in Leadership Studies?⁸

The above discussion described the development of a student leadership assessment instrument that involved multiple stakeholders and contextual grounding in the history and mission of the student leadership development program for which it was used as an evaluation tool. One of the challenges in using surveys (and most other data collection methods) is that there is no benchmark to know whether the assessed levels of leadership are acceptable or show a causal relationship to an intervention such as a leadership development program. Placing value or merit on the data collected with this or other student leadership assessment measures involves comparing the data with a standard. This valuing is the second arm of evaluation.

Summary

This article focused on two groups of higher education leaders, department chairs and students. First, it described the Individual Development and Educational Assessment (IDEA) Feedback for Department Chairs system at Kansas State University and its use to evaluate the effectiveness of department chairs across campus. Next, it presented the process used to develop a contextually based leadership assessment instrument for students in the university's Leadership Studies program. The recognition and development of leadership talent throughout institutions of higher education is a strategic imperative (Hill, 2005). The growing demand for accountability in higher education, the increase in emphasis on leadership at all levels, and the rapidly expanding number of programs and degrees in student leadership demonstrate the intersection of the fields of evaluation and leadership. Defining and assessing leadership qualities and competencies of department chairs and students, who may well become future leaders, is essential. As leaders in higher education, department chairs must exhibit top-notch professional competencies as well as conceptual and human competencies associated with leadership. Consideration of domains of leadership and expected indicators of successful leadership are contextually bound. Therefore, assessment instruments must consider context, content validity, and other important parameters of data collection methodologies. The use of appropriate evaluation designs or accepted standards is critical to evaluating leadership in higher education.

References

- Behrens, R., & Benham, M.K.P. (2007). Evaluating community leadership programs. In K.M. Hannum, J.W. Martineau, & C. Reinelt (Eds.), *The handbook of leadership development evaluation* (pp. 284-315). San Francisco: Jossey-Bass.
- Binard, K., & Brungardt, C. (1997). Learning leadership: Assessing students at the Community College of Denver. *The Journal of Leadership Studies*, 4(4), 128-104.
- Bogden, R.C., & Biklen, S.K. (1982). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn & Bacon.
- Brungardt, C., & Crawford, C.B. (1996). A comprehensive approach to assessing leadership students and programs: Preliminary findings. *The Journal of Leadership Studies*, 3(1), 37-48.
- Chambers, T. (1992) The development of criteria to evaluate college student leadership programs: A Delphi approach. *Journal of College Student Development*, 33, 339-347.
- Craig, S.B., & Hannum, K.M. (2007). Experimental and quasi-experimental evaluations. In K.M. Hannum, J.W. Martineau, & C. Reinelt (Eds.), *The handbook of leadership development evaluation* (pp. 19-48). San Francisco: Jossey-Bass.
- Cress, C.M, Astin, H.S., Zimmerman-Oster, K., & Burkhardt, J.C. (2001). Developmental outcomes of college students' involvement in leadership activities. *Journal of College Student Development*, 42, 15-27.
- Dillman, D.A. 2007. *Mail and internet surveys: The tailored design method* (3rd ed.). New York: Wiley.
- Doyle, J.K. (2008). *Introduction to survey methodology and design*. Retrieved March 12, 2008, from http://www.sysurvey.com/tips/introduction_to_survey.htm.
- Durland, M.M., & Fredricks, K.A. (Eds.). (2006). Social network analysis in program evaluation. *New Directions for Evaluation*, 107.
- Gmelch, W.H., & Miskin, V.D. (1993). *Leadership skills for department chairs*. Bolton, MA: Anker.
- Gmelch, W.H., & Miskin, V.D. (2007). Leading through teams. *Effective Practices for Academic Leaders*, 2, 1-15.
- Grove, J.T., Kiber, B.M., & Hass, T. (2005). *EvalLEAD: A guide for shaping and evaluating leadership development programs*. Oakland, CA: Sustainable Leadership Initiative/The Public Health Institute.
- Hannum, K.M., Martineau, J.W., & Reinelt, C. (Eds.). (2007). *The handbook of leadership development evaluation*. San Francisco: Wiley.
- Hill, L. (2005). Leadership development: A strategic imperative for higher education. *Forum for the future of higher education*. Retrieved January 21, 2009, from <http://connect.educause.edu/Library/Abstract/LeadershipDevelopmentAstr41253>.
- Hoyt, D.P., Bailey, M., Pallett, W.H., & Gross, A.B. (1999). *The IDEA system for evaluating and improving the administrative performance of department chairpersons: Revising the DECAD FORM*. Manhattan, KS: Kansas State University, IDEA Center.
- Kline, P. (1999). *The handbook of psychological testing* (2d ed.). London: Routledge.
- Kouzes, J.M. & Posner, B.Z. (1988). *The leadership practices inventory*. San Diego: Pfeiffer and Company.
- Leaming, D. R. (2007). *Academic leadership: A practical guide to chairing the department* (2d ed.). Bolton, MA: Anker Pub. Co.
- McDade, S.A. (1987). *Higher education leadership: Enhancing skills through professional development programs* (Vol. 5). Washington, DC: Clearinghouse on Higher Education.
- Middendorf, B.J., Benton, S., & Webster, R. (2009). *Using faculty ratings to provide formative feedback about the department chair's effectiveness*. Paper presented at Academic Chairpersons' Conference, Orlando, FL.
- Miller T. K. (Ed.). (2003). *The book of professional standards for higher education*. Washington, DC: Council for the Advancement of Standards.
- Mitchell, J.N. (2004). Measuring the performance of the chair. *New Directions for Higher Education*, 126, 55-68.
- Montez, J. (2003, April). *Developing an instrument to assess higher education leadership*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. ERIC Document Reproduction Service No. 477446.
- Peters, L., & Baum, J. (2007). The importance of local context in leadership development and evaluation. In K.M. Hannum, J.W. Martineau, & C. Reinelt (Eds.), *The handbook of leadership development evaluation* (pp. 261-283). San Francisco: Jossey-Bass.
- Poister, T.H. (2003). *Measuring performance in public and nonprofit organizations*. San Francisco: Jossey-Bass.
- Posner, B.Z., & Kouzes, J.M. (1992). *Psychometric properties of the leadership practice inventory*. San Diego, CA: Pfeiffer and Company.
- Ruben, B.D. (Ed.). (2004). *Pursuing excellence in higher education: Eight fundamental challenges*. San Francisco: Jossey-Bass.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Stufflebeam D. (1999). *Foundational models for 21st century program evaluation*. Kalamazoo, MI: Western Michigan University, Evaluation Center.
- Umble, K.E. (2007). Evaluation for planning and improving leadership development programs: A framework based on the Baldrige education criteria for performance excellence. In K.M. Hannum, J.W. Martineau, & C. Reinelt (Eds.), *The handbook of leadership development evaluation* (pp. 464-487). San Francisco: Jossey-Bass.
- Wheeler, D., Seagren, A.T., Becker, L.W., Kinley, E., Mlinek, D., & Robson, K.J. (2008). *The academic chair's handbook* (2d ed.). San Francisco: Jossey-Bass.

Williams, J.F. (2001). Leadership evaluation and assessment. *Journal of Library Administration*, 32 (3/4), 145-167.

Wisniewski, M.A. (1999). Leadership competencies in continuing higher education: Implications for leadership education. *Journal of Continuing Higher Education*, 7(1), 14-23.

Zimmerman-Oster, K., & Burkhart, J. (1999). *Leadership in the making: Impact and insights from leadership development programs in U.S. colleges and universities*. Battle Creek, MI: W.K. Kellogg Foundation.

Endnotes

¹ Double-barreled survey questions ask the respondent to assess two concepts in the same question. It is a problem with survey development.

² The FPDHS and CIF are found at <http://www.theideacenter.org/node/8>.

³ A sample is found at <http://www.theideacenter.org/sites/default/files/DeptChairSam.pdf>.

⁴ A sample is found at <http://www.theideacenter.org/sites/default/files/ChairSurveySample.pdf>.

⁵ Samples of these instruments are may be found at <http://www.theideacenter.org/node/8>.

⁶ Another method of placing value on assessed leadership is utilizing evaluation designs, including experimental or quasi-experimental research designs, to place value on leadership assessment data (Craig & Hannum, 2007; Shadish, Cook, & Campbell, 2002). Ongoing measures, such as those used at intervals during a leadership development program, lend themselves to time-series evaluation designs. Other possible designs are utilizing peer group comparisons or control groups. In addition to evaluating individual leadership, evaluation of collective leadership includes such methods as social network analysis (Durland & Fredericks, 2006) and ethnography (Behrens & Benham, 2007). Binard and Brungardt (1997) used a pre-post design to evaluate the impact of student leadership development activities. Customized open-systems frameworks were used to evaluate EvaluLEAD youth leadership programs (Grove, Kiber, & Hass, 2005), and the National Public Health Leadership Institute used the Baldrige Education Criteria for Performance Excellence Framework (Umble, 2007).

⁷ Items for the alumni survey were developed by referring to published literature related to expected outcomes of leadership studies programs (e.g. Cress, Astin, Zimmerman-Oster, & Burkhardt, 2001; Williams, 2001; Chambers, 1992) and program outcome data about program outcome expectations provided by various Leadership Studies program stakeholders (i.e., students, faculty/staff, founders). Expected outcomes were gathered via surveys of the program's advisory council; focus groups; students at various levels within the program; and program faculty and staff. Other sources used to inform the development of the instrument included historical documents provided by program faculty; information gathered during oral history interviews with the founders; and literature discussing various indicators of successful leadership. Semi-structured founder interviews followed a protocol that aimed to assist the participants in thinking

about historical events and experiences related to the founding of the program. In addition, they were asked about the students outcomes they expected to be produced by the program. The first level of analysis of the interviews regarded the program's expected impact on program participants. For coding purposes, expectations were defined broadly, inclusive of "must" and "should" (i.e., recommendations). The results of this analysis were used to inform the development of the survey for the advisory group and the program faculty. In developing the advisory group survey, a select group of council members who represented various program stakeholder groups (alumni, parents of alumni, employers of alumni, and business and civic leaders) were interviewed. Interview questions were created based on the information collected during a review of the program's historical documents and founders' interviews. Interview questions addressed what the council members saw as benefits to various stakeholders. Responses from the interviews as well as the oral history interviews were used to shape the questions included in the survey administered to the entire advisory group. Survey questions included requests to describe the benefits of the leadership program to students and alumni. The faculty survey was a modified version of the survey given to the advisory group.

⁸ The responses to the first set of questions were combined for all four groups and analyzed by theme using a qualitative approach (Bogden & Biklen, 1982). These were incorporated with other stakeholder input and sources of information related to indicators to develop the alumni survey. See Appendix for further detail.

Appendix

Development of Alumni Survey

All elements of this instrument development complied with the University's Institutional Review Board process. The development and implementation of these surveys conformed to Dillman's (2007) methodology recommendations for survey development and administration. During the survey development phase of the project, care was taken to use strategies to reduce non-response error and measurement error. The Dillman Tailored Design Method (TDM) is the standard methodology used for designing questionnaires. Providing social validation, avoiding subordinating language, making the questionnaire interesting, minimizing requests for personal information, and making the task important are recommended ways of developing trust (social exchange) within the framework on the questionnaire.

To minimize errors in the Leadership Studies Program Alumni Survey, Thurston and her team used Dillman's recommendations for wording questions, designing questionnaires, and pretesting the survey. The questionnaire was written in such a way that the questions were valid (that is, the questions measured what the researcher intended them to measure), reliable (the questions would yield the same results if administered at different times or to different samples), and unbiased (the questions were written in such a way that people would be willing and able to provide accurate answers). According to Doyle (2008), there are literally dozens of issues related to the precise wording of questions that should be carefully considered when constructing a survey. Thus, he suggested that all survey questions should be put through a "debugging procedure" in which several quality control questions are asked:

1. Is the question one that respondents can easily answer based on their experience?
2. Is the question simple enough, specific enough, and sufficiently well-defined that all of the respondents will interpret it in the same way?
3. Does the question contain any words or phrases that could bias respondents to answer one way over another?
4. Is it clear to respondents exactly what types of answers are appropriate?
5. Does the question focus on a single topic or does it contain multiple topics that should be broken up into multiple questions?
6. Are any listed response options mutually exclusive?

This process of writing, debugging, and revising survey questions was inherent in constructing the alumni survey. The process included repeated debugging and pretesting. The pretest included:

1. Reviewing the questionnaire by knowledgeable colleagues and analysts to obtain feedback about the substantive content of the questionnaire/ wording of questions, design of the survey, and validity of the content and questions.
2. Interviews to evaluate cognitive and motivational questions to answer such questions as: Are all the words understood? Are respondents likely to read and answer each question? Are all the questions interpreted similarly by all respondents? This step was combined with the previous step and with the next step, the pilot study.
3. Conducting a small pilot study using procedures that emulate the main study.
4. Conducting a final check by asking novice readers to double check for spelling and layout.

Using the Dillman (2007) steps for pretesting an instrument, the alumni survey was sent to an expert review panel composed of Leadership Studies Program faculty and staff. Revisions were made to the instrument based on reviewer feedback. To ensure clarity of the instrument, the revised survey was then pilot tested on a sample ($n = 30$) of 2008 alumni, who were not included in the final data collection. Eight alumni provided feedback, and revisions were made. The final alumni survey instrument consisted of scaled and open-ended items as well as demographic questions. The scaled items were developed to measure the fulfillment of each aspect of the Leadership Studies Program mission statement and the extent to which alumni agreed that participating in the program assisted them in achieving outcomes such as enhanced skills and abilities.