Effective Leadership in the Assessment of Learning Outcomes in Higher Education

Stephen L. Benton, Dan Li

IDEA Education

Rodney K. Goodyear

School of Education, University of Redlands

Paper presented at the annual meeting of the American Psychological Association, Washington, D. C., August 2014. Correspondence concerning this article should be addressed to Stephen Benton, The IDEA Center, 301 S. 4th St., Suite 200, Manhattan, KS 66502-6209. E-mail: steve@IDEAedu.org
Effective Leadership in the Assessment of Learning Outcomes in Higher Education

Much has been written about department chair roles and responsibilities, both in the U.S. and abroad (e.g., Braskamp, 2008; Buller, 2006; Chu, 2006; Creswell et al., 1990; Christie, 2007; Gmelch & Miskin, 1995; Hecht et al., 1999; Higgerson, 1996; Knight & Trowler, 2000; Middendorf & Benton, 2009; Seagren et al. 1993; see also http://www.departmentchairs.org/Books.aspx for a more comprehensive list of book titles). As leaders of their academic units, one of department chairs’ major responsibilities is to coordinate departmental priorities with the broader mission of the institution (Hecht et al. 1999).

External pressure from professional accreditors has made one of those priorities increasingly important: using data to improve teaching and to provide insight into student learning (Kezar, 2013; Kuh & Ikenberry, 2009; Vokwein et al., 2006). But whereas the impetus to do so may come from external pressures, the extent to which outcome assessment is successful depends heavily on such internal characteristics as leadership and faculty involvement (Peterson & Augustine, 2000). Unit leaders who are committed to transforming the organizational structure to support assessment are more successful carrying out meaningful assessment of learning outcomes (Lind & McDonald, 2003). The current study investigated which leadership qualities of department chairs are most strongly associated with successful performance of this responsibility. We also examined whether institutional enrollment size might moderate the relationships between those variables.

Our data source was the IDEA Education’s Feedback System for Chairs (FSC) (http://ideaedu.org/services/department-chairs). The system is comprised of both chair and faculty surveys. Chairs rate the priority of 21 responsibilities and self-assess personal characteristics and administrative methods associated with successful implementation. Faculty
then rate the chair on the same responsibilities, personal characteristics, and administrative methods.

Theoretical Framework

Thrust upon the shoulders of higher education administrators is the recognition by outside constituencies that outcomes mean more than inputs. For decades, it was assumed that college students were receiving a good education not by what they knew or could do but by the quality of faculty, the test scores of entering students, the size of library holdings, and other inputs. Now leaders in colleges and universities increasingly are required to show credible evidence of student learning outcomes and the investments that make a difference (Kuh & Ickenberg, 2009).

Department chairs are most typically the persons responsible for showing what students know and can do, though it is an assignment for which they rarely have received formal preparation. In this study, we wanted to identify administrative methods and personal characteristics of Chairs that are most associated with perceived success in planning, carrying out, and evaluating the assessment of student learning in a department. Few empirical studies have examined the effects of leadership on outcome assessment. Characteristics associated with successful implementation of outcomes assessment include trust (Braskamp, 1991), collegiality, and openness to risk (Banta, et al., 1996; Maki, 2010). What is needed, however, is research investigating key elements of leadership that make a difference. What research exists has consisted almost exclusively of single case studies that are merely descriptive. Moreover, most studies have employed the administrator’s perspective rather than the perceptions of those most affected by the administrator’s leadership -- faculty. What is needed are studies that examine the perceptions of a broader group of individuals who can help identify important leadership characteristics and behaviors (Kezar, 2013).
A review of the research on leadership and the assessment of learning outcomes revealed most studies have been conducted in the K-12 system. Assessment of student learning is not new to either K-12 or the business world (Kazin & Payne, 2009). But, in higher education, where autonomy and academic freedom are of supreme value, faculty engagement in assessment has been slow (Kuh & Ickenberry, 2009).

We, therefore, drew upon research concerning the impact of leadership dimensions on student learning outcomes in the K-12 setting. Although qualitative case studies provide numerous examples of strong principals who have made a difference in how much students learn in schools, quantitative studies have revealed equivocal findings. Some have found that school leaders have only minimal and indirect effects—mediated through teachers—on student learning outcomes (Hallinger & Heck, 1998). Others, as Marzano and colleagues (Marzano, Waters, & McNulty, 2005) reported in their meta-analysis of studies in the U.S., have found that principals have moderately strong effects.

Based upon a meta-analysis of international literature, Robinson (2007) proposed a model of leadership dimensions associated with positive impacts on K-12 student learning outcomes. She included in her search studies that investigated the effects of superintendents, principals, teachers, parents, and total school-based leadership on student learning in elementary, middle, and high schools. Outcomes were wide ranging but predominately included reading, mathematics, and language. Her search identified 26 studies published between 1978 and 2006, 18 of which were conducted in the U.S. From her meta-analysis Robinson (2007) identified five dimensions of leadership that have been empirically shown to affect K-12 student learning outcomes: establishing goals and expectations; strategic resourcing; planning, coordinating, and evaluating teaching and the curriculum; promoting and participating in teacher learning and
development; and ensuring an orderly and supportive environment. Their hypothesized relevance to university chairs’ leadership is described in the following paragraphs.

*Establishing goals and expectations.* In setting goals and expectations, chairs can indirectly affect student-learning assessment by focusing on and coordinating the work of faculty and staff. Leaders of K-12 schools with high achieving students emphasize the importance of communicating their goals and expectations (Heck, Larsen, & Marcoulides, 1990; Heck, Marcoulides, & Lang, 1991). Goal setting creates a “discrepancy between what is currently happening and some desired future state” (Robinson, 2007, p. 10). Staff and faculty are motivated to reduce the difference between the present state and the goal state. They focus on influencing and measuring student outcomes; consequently, their teaching has a renewed sense of purpose beyond simply preparing for class, delivering a lecture, and administering tests. We hypothesized that faculty perceptions of the chair’s performance in establishing goals and expectations would be positively related to perceived success in assessing student-learning outcomes. Specific chair behaviors included establishing and monitoring goals, assisting faculty with individual goals, maintaining performance standards, and seeing that faculty are working up to capacity.

*Strategic resourcing.* Department chairs can influence student learning through the actions they take concerning faculty recruitment/retention and the securing and allocating of resources. To have a positive effect on outcomes assessment, the resources must be assigned directly to support teaching and learning (Robinson, 2007). It is not sufficient, then, to be a good fundraiser; the money must be used to recruit high-quality instructors and to support professional development. In the current study, we identified a single-item indicator of strategic resourcing:
facilitating successful recruitment of quality faculty. We hypothesized that ratings on that item would be positively related to assessment of learning outcomes.

Planning, coordinating, and evaluating teaching and the curriculum. At the K-12 level, the more actively involved the principal is with teachers in overseeing and coordinating the instructional program, the greater the student progress. Principals who ensure faculty systematically monitor student-learning outcomes lead higher performing schools. Students make greater progress when their learning experiences are interconnected and coherent (Bransford, Brown, & Cocking, 2000). We hypothesized that chairs whose faculty perceive being active in evaluating teaching and facilitating curriculum development also would be perceived as having a greater impact on assessment of learning outcomes.

Promoting and participating in teacher learning and development. Effective leaders not only promote professional development for their faculty but also actively participate in it and often serve as a source of professional advice. Such involvement communicates a clear focus on the improvement of teaching and student learning and keeps chairs abreast about resource needs. Close contact with faculty helps them understand what needs to happen to bring about change. However, no items in the IDEA FSC were considered good measures of this leadership dimension.

Ensuring an orderly and supportive environment. Effective leaders foster a safe and secure environment so that effective teaching and high student achievement can occur. They communicate care and respect for all, foster multicultural tolerance and understanding, maintain an orderly environment with discipline, and allow minimal interruptions of teaching time. They build trusting relationships by caring about faculty, parents, and students. Consistency is shown in the congruence between what the leader says and does (Robinson, 2007). In this study, we
therefore hypothesized faculty ratings on the following characteristics would be positively related to the chair’s perceived success in assessing student learning outcomes: demonstrating caring, being trustworthy and consistent, resolving conflicts, looking out for personal welfare of faculty, and promoting inclusiveness among students and faculty.

Purpose of the Study

This study examined whether aspects of Robinson’s (2007) leadership dimensions—developed at the K-12 level—could be applied to explaining department chair success in implementing meaningful assessment of learning outcomes in higher education. We also examined whether institutional enrollment would moderate aspects of Robinsons’ (2007) leadership dimensions in explaining variance in faculty ratings of the chair’s performance.

Methods

Instrumentation

The instruments came from IDEA Education’s FSC. IDEA Education is a “nonprofit organization whose mission is to provide assessment and feedback systems to improve learning in higher education” (see http://ideaedu.org/v/about). IDEA Education supports the evaluation and development of a number of programs that focus on students, faculty, department chairs, deans, and other higher education administrators. In the FSC, department chairs complete the Chair Self-Assessment (CSA) form, and their corresponding faculty members respond to questions on the Faculty Perceptions of Chair (FPC).

On the CSA, chairs rate the importance of 21 administrative responsibilities, using the scale, 1 (Low Priority), 2 (Moderate Priority), or 3 (High Priority). Chairs rate themselves on 11
personal characteristics and 21 administrative methods, using a Likert scale from 1 (*Definite Weakness*) to 5 (*Definite Strength*). ¹

On the FPC, faculty rate their respective department chair’s performance of the same 21 administrative responsibilities by responding: 1 (*Poor*), 2 (*Fair*), 3 (*In Between*), 4 (*Good*), or 5 (*Outstanding*). They then rate the chair on the 11 personal characteristics and 21 administrative methods, using the same scale as on the CSA.²

**Data Source**

Data were accessed from IDEA Education’s FSC database of chair self-assessment and faculty ratings. Several exclusion criteria were enacted. First, the chair’s entry was removed if fewer than eight faculty members rated a specific department chair in a particular year. This criterion was based on previous research that found ratings were more reliable when at least 8 raters responded (Hoyt et al. 1999). Second, if a department chair was rated in multiple years, only one year of data was retained by random selection to maintain independence of responses. Third, only cases where a faculty member responded to at least 50% of the items on the FPC were retained to ensure the representativeness of a chair’s rating within a department. Data were aggregated for each chair, and mean faculty ratings were computed on each FPC item.

From 2010 to 2013, a total of 165 department chairs from 19 institutions in the U.S. completed the CSA. Corresponding faculty members (*N* = 2,713) completed at least 50% of the items on the FPC. Department size ranged from 8 to 53 faculty with the mode being 10. Institution enrollments ranged from 2,881 to 49,827 (*M* = 19,107, *SD* = 8,239). Faculty response rates typically average approximately 76% (Benton, Gross, Pallett, Song, and Webster, 2010).

¹ Samples of the Chair Self-Assessment form can be found at http://www.ideaedu.org/sites/default/files/sample_csa_survey.pdf
² Samples of the Faculty Perceptions of Chair form can be found at http://www.ideaedu.org/sites/default/files/sample_fpc_survey.pdf
Table 1 presents frequencies and percentages of various demographic characteristics for participating departments and institutions. Most chairs came from public institutions and from either the Midwest or Southeast. The vast majority represented doctorate-granting departments. Overall, although our exclusion criteria may have removed chairs and faculty with certain demographic characteristics (particularly those in smaller departments and smaller universities), it appears that we obtained a sample of chairs from a broad range of American universities.

Evidence of validity. Benton et al., (2010) present detailed evidence of reliability and validity. The initial development of the items for the CSA and FPC began with McCarthy (1972), who sought to address the lack of evaluation measures available for assessing academic department chairs. McCarthy (1972) reviewed relevant literature at the time, including several contemporary books written for or by department chairs. Based on that review, Hoyt and colleagues (Hoyt et al., 1999) created the initial 20 priorities, each assigned to one of 5 subscales: Administrative Support (communicates department’s needs, guides organizational plans, attends to administrative details); Personnel Management (guides faculty evaluation, leads faculty recruiting, fosters faculty development, rewards faculty); Program Leadership/Support (fosters good teaching, encourages balanced faculty, stimulates research/scholarly work, communicates administrative expectations, guides curriculum development, leads department planning); Building Image/Reputation (facilitates external funding, improves on-campus image, improves off-campus image); and Developing Positive Climate (fosters collegiality, orients new faculty/staff, stimulates faculty vitality, establishes trust). Chair ratings of the importance of these responsibilities have remained consistent across the years (Middendorf & Benton, 2009).

Staff at IDEA Education updated the CSA and FPC in 2010 based on extensive statistical analyses of the chair and faculty ratings database, input from four focus groups, and an external
expert review panel. The focus group participants included faculty, department chairs, and other administrators from colleges and universities ranging in enrollments of 3,000 to 30,000. It also included individuals from two- and four-year schools and public and private institutions. The panel was comprised of current and former academic administrators, some of whom had written highly regarded books on chairing the academic department. Based on these sources, 9 items were dropped from the previous FPC, 24 were reworded, and two new administrative responsibilities were added (Benton et al., 2010).

Benton and colleagues (2010) found evidence for four underlying dimensions of responsibilities, using principal components analysis with varimax rotation. The first factor, Personnel Management and Development (6 items; Cronbach’s $\alpha = .81$), concerned facilitating research/grants and as well as appropriately recruiting, assessing, and rewarding faculty. The second factor, Developing Positive Climate (4 items; Cronbach’s $\alpha = .84$), dealt with fostering faculty talents and interests, rejuvenating faculty vitality/enthusiasm, and developing collegiality. Academic Support/Program Leadership (5 items; Cronbach’s $\alpha = .76$) addressed attending to administrative details, communicating expectations of the campus administration to the faculty, and acquainting new faculty with departmental procedures. Lastly, Building Image (2 items; Cronbach’s $\alpha = .77$), concerned fostering a good image of the department on- and off-campus.

Benton and colleagues (2010) also offered evidence for the validity generalization of chair and faculty ratings across several demographic variables: Carnegie classification, public versus private institutions, chair years of experience, and highest degree awarded in the department. The high degree of consistency across these demographics in ratings of the 20 priorities demonstrated their relevance for diverse types of institutions, faculty, and chair experiences. Additional evidence was provided for criterion-related validity by correlating chair
and faculty mean ratings on the administrative responsibilities. The Pearson $r$ was significant ($r = .69, p = .001$), which indicated a moderately high relationship existed between chairs’ ratings of importance and faculty ratings of performance (Benton et al. 2010). Because chairs and faculty were rating different constructs (importance vs. performance), and using different response scales, one would not have expected such a high correlation. Nonetheless, chairs tended to perform better—at least in the eyes of faculty members—on responsibilities they considered important. In contrast, chairs received lower ratings on responsibilities they considered less important. This is consistent with the underlying purpose of the instrument.

**Evidence of reliability.** Benton et al. (2010) provided reliability evidence through the split-half method by correlating item scores from subgroups of faculty rating the same chair. They first selected chairs for which the number of faculty raters equaled 12 or more. When there were more than 12 raters, 12 were randomly selected. They then performed 1,000 random splits ($n = 6$ each) on each chair’s raters, computed means for each split, correlated the subgroup means, and computed the mean $r$ for each item. Using the Spearman-Brown prophecy formula, they then estimated the reliability of ratings for department sizes of 8, 12, 16, and 20 faculty. For departments of at least 8 raters the reliability coefficients ranged from .52 to .72; for 12 raters from .62 to .80; for 16 raters from .64 to .84; and for 20 raters from .73 to .87.

**Results**

Correlations among all variables are presented in Table 2. Separate moderated hierarchical regression models were computed, using SPSS, for four of the five hypothesized dimensions of leadership. All variables were standardized prior to entry. For all models, enrollment was entered on the first step to control for institution size. On the second step, main effects for each explanatory variable were entered followed on the third step by all possible two-
way interactions with enrollment. In all cases, the dependent variable was mean ratings of the chair’s success at ensuring meaningful and ongoing assessment of student learning outcomes. Type I error rate was set at $\alpha = .05$.

We began by first testing the model containing explanatory variables conceptually related to establishing goals and expectations (see Table 3). On the first step, enrollment had a small negative influence ($\beta = -.16, p < .053$) on the dependent measure ($R^2 = .018$). Ratings on the criterion variable tended to be somewhat lower as enrollments increased. On the second step, adding the main effects from the explanatory variables contributed an additional 71% of variance ($R^2 \text{ Change} = .710$). Establishing and monitoring progress on department goals ($\beta = .28, p < .004$), assisting faculty in developing their own goals ($\beta = .21, p < .025$), and seeing that faculty work up to capacity ($\beta = .27, p < .015$) each contributed positively to the model. Ratings on the criterion increased, then, as faculty perceptions of the chair’s administrative methods increased.

On the third step, only the interaction of Enrollment by Assisting Faculty was significant ($\beta = -.27, p < .011$). We then plotted lines of best fit for ratings on “assisting faculty with goals” at three levels of enrollment (mean enrollment, +1 SD above mean enrollment, and -1 SD below mean enrollment). The plots showed an ordinal interaction: The superior performance of chairs at institutions with low enrollments decreased somewhat as success in assisting with faculty goals increased (see Figure 1). So, helping faculty with goal setting was especially beneficial in large institutions.

For strategic resourcing, after enrollment was entered on the first step, ratings on faculty recruitment ($\beta = .79, p < .001$) added significantly to the model on the second step ($R^2 \text{ Change} = .613$). Chairs who were perceived as doing a good job of facilitating recruitment of promising
faculty earned higher ratings on assessment of learning outcomes. However, there was no significant interaction between enrollment and faculty recruitment (see Table 4).

In the model testing variables related to planning (see Table 5), coordinating, and evaluating teaching and the curriculum, on the second step all three explanatory variables made significant contributions: fostering good teaching ($\beta = .21, p < .012$), assessing faculty performance ($\beta = .30, p < .001$), and facilitating curriculum development ($\beta = .44, p < .001$). The additional explained variance on the second step was substantial ($R^2 = .813$). Again, ratings on the dependent measure were higher when chairs were perceived as having strengths in key administrative methods. There were no significant two-way interactions.

Finally, the model for testing the leadership dimension of ensuring an orderly, supportive environment revealed several significant relationships (see Table 6). After entering enrollment on the first step, main effects were observed for consistency ($\beta = .37, p < .001$) and reducing conflict ($\beta = .37, p < .001$). Ratings on assessment of learning outcomes improved as perceptions of the chair’s consistency and conflict resolution skills increased. On the third step, four two-way interactions reached significance: Enrollment x Caring (see Figure 2), Enrollment x Conflict Resolution (see Figure 3), Enrollment x Faculty Welfare (see Figure 4), and Enrollment x Promotion of Inclusiveness (see Figure 5). The plots of lines of best fit followed the same pattern described previously: The superior performance of chairs at institutions with low enrollments decreased somewhat as success in ensuring an orderly and supportive environment increased. So, again, chairs are large institutions gained the most from ensuring order and support.

Conclusions
Several aspects of Robinson’s (2007) leadership dimensions are associated with perceived chair success in implementing meaningful assessment of student learning outcomes. In *establishing goals and expectations*, important chair behaviors include establishing and monitoring goals, assisting faculty in developing their own goals, and seeing that faculty work up to capacity. Enrollment size moderates the effect of the latter behavior as chairs in institutions with high enrollments earn higher ratings on student assessment when they hold faculty accountable for their work.

Concerning *strategic resourcing*, facilitating successful recruitment of promising faculty is positively associated with ratings of the chair’s performance. With respect to *planning, coordinating, and evaluating teaching and the curriculum*, it is important for the chair to foster good teaching, guide the development of sound procedures for assessing faculty performance, and facilitate curriculum development. Finally, to *ensure an orderly and supportive environment*, consistency and conflict resolution skills are of value. In institutions with larger enrollments, demonstrating caring, being skilled in conflict resolution, watching out for faculty welfare, and promoting inclusiveness are especially important for perceived success in assessing student-learning outcomes.

**Implications**

These findings provide some useful direction for current and aspiring chairs. They show that chairs, for example, are likely to find themselves judged by the extent to which they are successful in recruiting good faculty and coordinating and evaluating both the curriculum and the effectiveness of individual instructors.

Some of the results though are moderated by institutional size, which is a proxy for several interrelated factors that all bear on leader behavior. For example, size reasonably will
affect both the size of the departments and how they are organized; it also will affect the extent to which there are specialized personnel available to share the burden of program assessment; and it will affect the both the chair’s familiarity with and closeness to the departmental faculty.

Therefore, *establishing goals and expectations* is important for all chairs, but becomes the more so as department size increases. For example, it is possible that in a smaller department peer pressure will be more effective in helping to ensure faculty involvement in learning assessment, whereas in larger departments the chair is more responsible for ensuring accountability.

*Limitations*

We acknowledge several limitations. First, participation was voluntary, and the reasons chairs had for selecting the IDEA Feedback system varied. Some may have been required to participate, some may have wanted feedback for professional development, whereas others may simply have been interested in finding out how they were perceived. Nonetheless, the sample represented a widespread group of chairs from many regions of the country, and institutions of various sizes and affiliations. Second, ratings were based on individual perceptions of the respondent and not on the administrator’s actual performance. Other measures of achievement (e.g., external funding, program development, effective budgeting, direct measures of student outcomes) would have been helpful to more validly assess the chair’s effectiveness. Third, the results of this study were correlational and say nothing about which administrative behaviors and interpersonal characteristics might cause one to be a more effective leader. Fourth, we have presented no evidence that providing chairs with feedback will necessarily improve their performance or
their ratings. Future research should address this question in the hopes of identifying efficacious professional development for future higher education leaders.
References


Table 1

*Frequencies and Percentages of Demographic Categories for Participating Chairs (N = 165)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>f</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutional Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>152</td>
<td>92.1</td>
</tr>
<tr>
<td>Private</td>
<td>13</td>
<td>7.9</td>
</tr>
<tr>
<td>Region of U.S. -- Institution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>7</td>
<td>36.8</td>
</tr>
<tr>
<td>Southeast</td>
<td>6</td>
<td>31.6</td>
</tr>
<tr>
<td>East</td>
<td>3</td>
<td>15.8</td>
</tr>
<tr>
<td>West</td>
<td>2</td>
<td>10.5</td>
</tr>
<tr>
<td>North Central</td>
<td>1</td>
<td>5.3</td>
</tr>
<tr>
<td>Highest Degree Awarded by Department</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associate</td>
<td>7</td>
<td>4.2</td>
</tr>
<tr>
<td>Baccalaureate</td>
<td>4</td>
<td>2.4</td>
</tr>
<tr>
<td>Professional</td>
<td>2</td>
<td>1.2</td>
</tr>
<tr>
<td>Masters</td>
<td>11</td>
<td>6.7</td>
</tr>
<tr>
<td>Masters +</td>
<td>9</td>
<td>5.5</td>
</tr>
<tr>
<td>Doctorate</td>
<td>132</td>
<td>80.0</td>
</tr>
</tbody>
</table>