Developing a Multi-Dimensional Evaluation Framework for Faculty Teaching and Service Performance

Diane F. Baker, Ph.D.
Professor of Management
Else School of Management
Millsaps College
Jackson, Mississippi

Walter P. Neely, Ph.D.
Professor of Finance
Else School of Management
Millsaps College
Jackson, Mississippi

Penelope J. Prenshaw, Ph.D.
Professor of Marketing
Else School of Management
Millsaps College
Jackson, Mississippi

Patrick A. Taylor, Ph.D.
Associate Professor of Economics
Else School of Management
Millsaps College
Jackson, Mississippi

ABSTRACT

A task force was created in a small, AACSB-accredited business school to develop a more comprehensive set of standards for faculty performance. The task force relied heavily on faculty input to identify and describe key dimensions that capture effective teaching and service performance. The result is a multi-dimensional framework that will be used by faculty and administrators to communicate and assess performance expectations. The dimensions for evaluating teaching are 1) quality of instruction, 2) innovative practices, and 3) effort or time commitment. The service dimensions are 1) contribution & leadership and 2) team player. The paper describes the steps taken to develop the framework and build faculty support for its use. The experiences of the task force and the process they used to develop the framework are relevant to those who are interested in revising and clarifying the faculty evaluation process. The final version of the framework is provided in the appendix.

INTRODUCTION

How should faculty of our colleges and universities be evaluated? What criteria should evaluation committees, chairs and deans use to assess faculty performance? The faculty evaluation process within institutions of higher education is of upmost importance in determining progress towards tenure, promotion, and salary. Accrediting bodies such as The Southern Association of Colleges and Schools (SACS) and the Association to Advance Collegiate Schools of Business (AACSB) typically require faculty evaluation. Standard 3.7.2 in the SACS "Principles of Accreditation" (2011, 31) states, "The institution regularly evaluates the effectiveness of each faculty member in accord with published criteria, regardless of contractual or tenured status." AACSB requires that the faculty evaluation "process should extend beyond student evaluations of teaching and include expectations for continuous improvement" (34).

Despite its importance, faculty evaluation is viewed by many as unsatisfactory (Silva and Thomsen, 2013). Miller and Seldin (2014) concluded that "meaningful evaluation of faculty performance was rare and that judgments fre-

quently were based on information gathered in haphazard, even chaotic, fashion" (p. 35). Research related to performance evaluation is highly developed in the management literature, whereas in higher education performance review has received much less attention (Shepherd, Carley and Stuart, 2009).

Faculty dissatisfaction with the evaluation process was a concern for the dean of the Else School of Management at Millsaps College. The Else School of Management (the business school) is accredited by the AACSB and offers undergraduate, graduate, executive and certificate programs. Annual faculty evaluations begin with each faculty submitting a report documenting activities in the areas of teaching, research and service. The report is submitted to the dean of the business school for review, followed by the dean's response. There are no formal guidelines specifying content within the three academic areas. According to the college's Faculty Handbook (2007) tenure guidelines, faculty must demonstrate "sustained, noteworthy teaching," "serious attention to the duties and responsibilities of a faculty member" and "developed scholarship or noteworthy performance in the creative arts," (26) but no criteria for the annual evaluation process are identified.

Consistent with the findings of Silva & Thomsen (2013), the Else School of Management faculty expressed dissatisfaction with the evaluation process. In 2013, a survey of the business faculty indicated that only 23% percent were satisfied or very satisfied with how teaching was evaluated and 35% were satisfied or very satisfied with how service was evaluated. In a similar survey conducted 5 years earlier, dissatisfaction primarily involved the extent to which G.PA. and student ratings were used to assess faculty teaching performance.

In response to faculty concerns, the dean of the business school created a task force charged with identifying the standards that should be used to evaluate business faculty in teaching and service. She chose task force members from economics, management, finance and marketing to serve on the task force. While our business school is unique in that we are small and embedded in a liberal arts college, the concerns our faculty have about the evaluation process are not at all unique. As noted above, dissatisfaction with the faculty evaluation process is common across schools and disciplines. The steps taken by this task force to assess and revise the approach we use to evaluate faculty may be useful to other faculty and administrators who are also dissatisfied with the way in which faculty members are evaluated.

Laying the Groundwork

In our first meeting, the task force identified the fundamental objectives that the faculty evaluation process should achieve, and this informed the steps we took to develop a framework for faculty evaluation. According to Gabris & Ihrke (2001), performance appraisals are often used to: 1) provide feedback; 2) influence employee behavior; and 3) make merit decisions. The task force recognized that these three objectives were relevant for faculty evaluation. We wanted to develop a set of performance standards that would encourage clear communication among faculty and administrators about the behaviors and outcomes that support the college and business school missions and strategies. The framework had to be consistent with the college's current standards for promotion, tenure and merit pay decisions. Our intention was to create a tool that would help faculty clearly understand performance standards and have access to the feedback that could help achieve them.

The task force also recognized some of the challenges associated with faculty evaluation. Kluger & Densi (1996) found, for example, that feedback does not consistently increase employee performance. Their research indicated that increased performance is less likely when the focus of the feedback is on the person, as opposed to the tasks or the specific behaviors that will lead to improvement. In the case of faculty evaluations, an emphasis on "overall instructor rating," an item that currently appears on the Millsaps' student evaluation of teaching form, seems to be targeted toward the individual and may not lead to behaviors that improve teaching effectiveness. Lang & Kersting (2007) found that student ratings did not have an appreciable impact on teacher performance over time.

Another common problem relates to the validity of performance evaluations that do not use objective measures or outcomes. For example, in their review of the literature, Rynes, Gerhart & Parks (2005) noted that individual contributions are difficult to assess objectively when employees work in teams. This is likely to apply to assessment of faculty service, where much of the work is done in committees. Rynes et al. also indicated that interrater reliability is low for behavior-oriented assessment and the link between pay and performance is generally perceived as weak when such assessment is used. Low validity can be the result of poorly constructed evaluation instruments, rater biases, a rater's limited access to relevant performance information, among other problems. Results-based measures may be more objective, but tend to be deficient in that they fail to account for the full range of performance expected. Validity is likely to increase when evaluations of teaching and service performance is based on multiple

forms of evaluation (Shao et al. 2007; Stark-Wroblewski et al. 2007; Marsh and Roche 1997; Shields 1996).

The ideal performance evaluation process is both valid and perceived as fair by the employees. Perceptions of procedural justice, interactional justice and distributive justice have been found to affect employee reaction to performance appraisal, including responses related to morale, satisfaction, commitment to the organization and intentions to leave (Erdogan, 2002; Flint, 1999; Holbrook, 2002). Perceptions of procedural justice increase when employees are fully aware of the standards of performance, the standards are consistently applied and the employee has input into the process. Interactional justice is a function of the communication between the supervisor and the employee during the performance appraisal process. Distributive justice involves the perceived fairness of the outcomes associated with the performance evaluation process, including ratings, commendations and pay increases.

The primary concern of the task force was on procedural justice. We wanted to develop a framework that faculty members would support, could consistently be applied and effectively used to provide faculty with clear, formative feedback. Interactional and distributive justice issues are largely up to the administrators who would use the tool we designed, but our objective was to offer a framework that would support their responsibilities in maintaining a just process. To facilitate higher levels of procedural justice, the task force determined that it was imperative to have faculty input at every stage of the design. Researchers have consistently found that employee voice is key in increasing perceptions of procedural justice (Erdogan, 2002; Flint, 1999; Holbrook, 2002).

In summary, the task force objectives were to create a set of standards that: 1) were consistent with the college's mission and strategies; 2) would clarify for faculty and administrators the performance required to be promoted and earn merit awards; and 3) were perceived as fair by faculty. Essentially, we were hoping to improve communication among faculty and administrators, to offer a common language that would give all faculty the opportunity to develop and succeed. In both teaching and service, we wanted to utilize a variety of information sources to enhance the validity of the process. Because the business school at Millsaps College is AACSB accredited, the criteria used to evaluate research productivity are more clearly defined and faculty fully understand the research expectations. The focus of the task force was therefore on assessment of teaching and service, not on research.

What Does Current Research Tell Us About Faculty Evaluation?

The research surrounding faculty evaluations of teaching service can be categorized into two primary streams. The first stream of research examines aspects related to overall faculty performance, and in particular the relative weights given to teaching, research and service for use in promotion, tenure, and salary decisions. The second stream consists of a vast amount of research investigating the use of student evaluations for assessing teaching performance. Both streams are useful in identifying the factors that should be considered in assessing faculty performance.

Teaching Performance

Most studies reported that student evaluations of teaching (SETs) are the primary source of information used to assess faculty teaching performance (Clayson, 2009; Honeycutt, Thelen and Ford 2010; Stark-Wroblewski, Ahlering, and Brill 2007; and Williams and Rhodes 2002). In his meta-analysis, Clayson (2009) found virtually all business schools use some form of student evaluation to assess teaching performance. Much of the research describes the problematic nature of SETs as tools for assessing teaching effectiveness, most questioning their validity and raising concerns about bias (Peterson, Berenson, Misra, and Radosevich 2008; Chonko 2006; Algozzine, Beattie, Bray, Flowers, Gretes, Howley, Mohanty, and Spooner 2004; Engelland 2004; Nasser and Fresko 2002; Stapleton and Murkison 2001; Simpson and Siguaw 2000; and Marsh and Roche 1997). In contrast, Wright & Jenkins-Guarnieri (2012) found that SETs were a valid measure of teaching effectiveness, as measured by student achievement on student final exams and grades. They recommended the use of SETs if used together with "constructive, consultative feedback" (p. 694).

While student evaluations of teaching are the most common tool utilized to assess teaching performance, studies investigating overall faculty performance provide insight into additional approaches used to evaluate teaching performance. Shepherd, Carley, and Stuart (2009) surveyed marketing chairs from AACSB-accredited institutions, including both doctoral-granting and non-doctoral-granting. All institutions reported using student evaluations to measure teaching effectiveness. Other common approaches, in order of popularity, were teaching portfolios, classroom observation, peer review of syllabi, class size, and delivery approach (online or face-to-face). The only significant difference found between the two different types of institutions was that doctoral-granting schools were more likely to consider the number of graduate classes taught.

Three other studies are helpful in identifying the factors that college administrators typically use to evaluate teaching performance. Honeycutt et al. (2010), asked marketing chairs from AACSB-accredited institutions to assign 100 points among various methods used to evaluate quality teaching. Responses among balanced, research-focused and teaching-focused institutions were compared. By far, SETs received the most points by all three types of institutions, followed by observation by peer faculty and unsolicited student comments. Williams and Rhodes (2002) surveyed chief academic officers at four-year colleges and universities and found among all Carnegie classifications student ratings were used to a greater extent than all other methods, followed by chair evaluation, dean evaluation, self-evaluation, and course syllabi/exams/handouts. Committee evaluation, colleague opinions and classroom visits received moderate use. Teaching portfolios, informal student opinions, student exam performance, long-term student follow-up and enrollment in elective courses were used less often in the evaluation of faculty teaching. Miller & Seldin's (2014) survey of deans from four year liberal arts colleges were consistent with Williams & Rhodes (2002). indicating that student ratings, chair evaluations, self-evaluations, classroom visits and committee evaluations were commonly used to evaluate faculty teaching. Over a ten-year period, Miller & Seldin's (2012) found that the use of self-evaluations and classroom observations has increased over the ten-year period between 2000 and 2010.

The above research consists of surveys from administrators reporting current approaches for evaluating teaching. Adding to this literature is a study by Shao, Anderson, and Newsome (2007) who included faculty in addition to administrators when surveying AACSB-accredited institutions. In addition to reporting current practice, the study also asked "what should be used" to evaluate teaching effectiveness. Among 20 general items, SETs ranked first in current practice, and ranked second among those items that should be considered. Respondents believed "current in field" should be the most important, but was only given moderate importance in current practice. Other items that were believed to have too little weight included: peer's evaluations, classroom visits, class assignments, and alumni feedback. And, respondents thought the following received too much weight: teaching awards, use of technology, colleagues' opinions, course level (graduate/undergraduate), course type (required/elective), and class enrollment.

It is clear from the research examined above that the primary means for evaluating teaching performance are SETs. Used to a lesser extent are administrators' evaluations, classroom observation, review of syllabi and faculty self-evaluations.

Service Performance

We found several studies that examined service performance criteria. In the Honeycutt et al. (2010) study, responses were categorized based on institutional importance given to teaching and research. Among institutions that focused on teaching or have a balanced focus on teaching and research, the following activities were identified for evaluating service performance (in order of importance): (1) service to school/college; (2) service to department; (3) service to university; (4) service to discipline; and (5) service to business community. Williams and Rhodes (2002) surveyed chief academic officers at four-year colleges and universities and found among private institutions with bachelor's Carnegie classification the following service activities in order of importance: (1) service on college-wide committee; (2) academic advising; (3) service on departmental committee; (4) department administrative duties; and (5) advisor to student organizations.

Examination of studies involving evaluation of overall faculty performance offers additional insights for evaluating service performance. In Miller and Seldin (2014), deans from accredited four-year liberal arts colleges weighed activities considered 'a major factor' for promotion. Over 70% of deans cited campus committee work as a major factor, an increase from 58% in a 2000 survey. Student advising was cited almost as often as a major factor. The deans were much less likely to rate public service, activities in professional societies and consultation as a major factor in promotion decisions.

More limited in geographic scope, Cipriano and Riccardi's study (2005) compared the perceptions of 917 faculty and 118 department chairs from the Connecticut State University system. Respondents were asked to indicate factors they considered important in making personnel decisions. Department committee work was considered a major factor by most faculty (61%) and chairs (66%). Less than half of both faculty and chairs viewed campus committee work as a major factor in personnel decisions. More dissonance between faculty and chairs was found among other service criteria. Only 28% of faculty weighed activity in professional services as important, whereas 41% of chairs cited it as major factor. Student advising received endorsement as a major factor by 39% of chairs and only 27% of faculty. Differences were also found for service to the community, with chairs at 24% and faculty at 15%, and public service with chairs at 21% and faculty at 10%. Lastly, few chairs and faculty weighed consultation as important, with only 5% considering it a major factor.

Based on the literature review, the most commonly used factors to assess faculty service performance are service to the department, service to the college and/or university, professional activities, service to the community, academic advising, and consulting. There are no studies describing or evaluating the way in which participation in these activities are assessed.

Summary

Most of the studies we found surveyed administrators and/or faculty to identify the factors that are typically considered when evaluating faculty performance in teaching and service. Across institutions of higher education and among faculty and administrators, there is considerable consistency about what those key factors are. To evaluate teaching performance, student evaluations are used by almost all surveyed institutions. To assess service performance, the factors that are often cited as important include service work on behalf of the institution, service to the discipline and academic advising. However, none of the studies we found provided specific performance criteria. For example, what are acceptable SET scores? What constitutes acceptable levels of performance on a college or professional committee? In addition, we could not find examples of a faculty-led effort to develop specific standards with which to evaluate faculty teaching and service. We hope this paper fills this void by offering direction as institutions grapple with issues surrounding the evaluation of teaching and service performance.

Initial Faculty Input

Our task force generated a list of items that were cited in previous research and should be considered when evaluating teaching and service performance. In addition to the items found in the literature, we considered rubrics employed by previous administrations, conversations with colleagues and the business school dean, past experience in tenure and promotion decisions, and the broad tenure and promotion guidelines found in our faculty handbook.

We designed a questionnaire in order to learn how colleagues felt about possible factors that might be used to evaluate faculty performance. We ultimately identified 23 items related to teaching and 11 related to service, and asked faculty to rate the importance of each. A 5-point Likert scale was employed using "very important," "somewhat important," "neutral," "somewhat unimportant," and "not at all important." The questionnaire was emailed to all 15 full-time faculty, excluding part-time instructors and adjuncts. Anonymous responses were collected from 14 of the 15 full-time faculty members.

The number of faculty rating each item as either "somewhat important" or "very important" for evaluating teaching and service performance were combined and the results are presented in Table 1. Note that "heavy service" is an imprecise term but is generally understood by the faculty members and administrators to mean serving on committees requiring frequent meetings for which some advanced preparations is required and which attend to crucial institutional matters.

The teaching performance item believed to be very or somewhat important by the greatest number of faculty was "Number of Separate Preparations" with 13 (93%) mentions. Items believed important by more than 11(85%) of the faculty were "Accessibility to Students," "Number of Required Courses Taught," and "Use of Innovative Teaching Practices." More than 10 (75%) of faculty also attributed importance to "Student Ratings of Teaching Behaviors" (e.g., the instructor specifies goals, is well prepared, gives clear direction, etc.), "Student Overall Rating of Instructor," "Class Size," and "Teaching a New or Significantly Revised Course." Items receiving more moderate support included "Supervision of Directed Studies/Internships," "Teaching in Graduate Program," "Number of Elective Courses Taught," "Research in Collaboration with Students," "Teaching College Core Course," "Teaching in Executive Program," "Community-Engaged Learning," "Evaluation of Syllabi," and "Advising of Honors Student." Perceived as lacking in importance by faculty include "Student Mentoring," "Plans for Teaching Improvement," "Teaching in International Program," "GPA Ranking," "Students Course Grade," and "Teaching in Summer School."

Among the activities to evaluate service performance, 13 (93%) of the faculty believed "Heavy Service to the Business School" was important, followed by "Heavy Service to the College" and "Other Service to the Business School." More moderate support was given to "Other Service to the College," "Service to the Profession," "Chaired Committees," "Contribution to Comprehensive Exams," and "Meeting Deadlines." Generating average support from the faculty were "Community Service," "Business School Administrative Duties," and "Consulting." Few faculty believed "Advisor to Student Organizations," or "Teaching in Certificate Program" were important for evaluating service performance.

Identifying Performance Dimensions

Relying heavily on the survey results, the task force proceeded with development of a multi-dimensional evaluation framework for both teaching and service performance. In reviewing those teaching items that received more than 10 (75%) faculty endorsements as important for evaluating teaching, commonality across 3 dimensions became apparent. A first dimension identified, Quality of Instruction, is indicated by the faculty's belief that both

Table 1
Evaluating teaching and service performance
Number of faculty members (14) indicating the item was either "somewhat important" or "very important" for evaluating teaching and service performance

Teaching Items	Number	Service Items	Number
Number of Separate Preparations	13	Heavy Service to the Business School	13
Accessibility to Students	12	Heavy Service to the College	12
Number of Required Courses Taught	12	Other Service to the Business School	12
Use of Innovative Teaching Practices	12	Other Service to the College	10
Student Ratings of Teacher Behaviors	11	Service to the Profession	10
Student Overall Rating of Instructor	11	Chaired Committees	9
Class size	11	Contribution to Comprehensive Exams	9
Teaching a New or Significantly Revised Course	11	Meeting Deadlines	9
Supervision of Directed Studies/Internships	10	Community Service	8
Teaching in Graduate Program	9	Business School Administrative Duties	8
Number of Elective Courses Taught	9	Consulting	8
Research in Collaboration with Students	9	Advisor to Student Organizations	6
Teaching College Core Course	8	Teaching in Certificate Programs	4
Teaching in Executive Program	8		
Community Engaged Learning	8		
Evaluation of Syllabi	7		
Advising an Honors Student	7		
Student Mentoring	6		
Plans for Teaching Improvement	6		
Teaching in International Program	6		
GPA Ranking	5		
Students' Course Grade	5		
Teaching in Summer School	4		

the Student Ratings of Teaching Behaviors and Student Overall Rating of Instructor should carry weight in assessing teaching performance. This is in line with the research discussed above regarding faculty belief in the importance of student evaluations for measuring teaching effectiveness. However, student evaluations were not the only, or perhaps even the most important, item in evaluating faculty according to the business faculty who completed our survey.

Two similar items endorsed by more than 10 (75%) faculty were Use of Innovative Teaching Practices and Teaching of New or Significantly Revised Course. Thus, a second dimension was titled Innovative Practices to capture items related to innovation. The inclusion of innovation as a dimension for evaluating teaching is affirmed by the fact that the AACSB emphasizes its importance in its accreditation standards (AACSB, 2013). The teaching performance results also indicate the importance of items not directly related to quality of instruction: Number of Separate Preparations, Accessibility to Students, Number of Required Courses Taught, and Class Size. Research has identified these items as extraneous or situational factors that influence student evaluation scores (Peterson et al. 2008; Engelland 2004; Wachtel 1998; and Marsh and Roche 1997). Given the importance of these items to our faculty, we encapsulated them into a third dimension identified as Effort or Time Commitment. See Table 2 for the breakdown of the dimensions for teaching performance.

Table 2
Dimensions of teaching performance: quality, innovation, and effort

Quality of Instruction	Number (%)
Student Ratings of Teacher Behaviors	11 (77)
Student Overall Rating of Instructor	11 (77)
Evaluation of Syllabi	7 (50)
Plans Teaching Improvement	6 (43)
GPA Ranking	5 (36)
Students' Course Grade	5 (36)
Innovative Practices	**Number (%)**
Use of Innovative Teaching Practices	12 (86)
Teaching a New or Significantly Revised Course	11 (77)
Research in Collaboration with Students	9 (64)
Community Engaged Learning	8 (57)
Teaching in International Program	6 (43)
Effort or Time Commitment	**Number (%)**
Number of Separate Preparations	13 (93)
Accessibility to Students	12 (86)
Number of Required Courses Taught	12 (86)
Class Size	11 (77)
Supervision of Directed Studies/ Internships	10 (71)
Teaching in Graduate Program	9 (64)
Number of Elective Courses Taught	9 (64)
Teaching College Core Course	8 (57)
Teaching in Executive Program	8 (57)
Advising an Honors Student	7 (50)
Student Mentoring	6 (43)
Teaching in Summer School	4 (28)

The service performance criteria clustered around 2 dimensions: Contribution & Leadership and Team Player. Discussion among task force members and among the faculty at large revealed that it was important to make contributions to the business school, college and profession, but it also mattered how well faculty worked with others to achieve outcomes. The Millsaps College handbook states, for example, that "serious attention to the duties and responsibilities of a faculty member [include] . . . evidence of cooperative interaction with colleagues, respect for the abilities of others, willingness to work toward a common purpose . . .". With only 15 full-time faculty in the Else School of Management, being a team player is critical and thus its perceived importance for evaluating service performance is warranted.

Once the five dimensions were defined, the next step in the refinement process was to establish expectations against which faculty are evaluated. The dean asked the task force to describe three levels of performance: meets expectations, exceeds expectations and does not meet expectations. The task force recognized that courses, committees and other responsibilities can vary considerably within a department, and faculty may have little control over variables such as committee assignments or class size. We agreed that it would be impossible to create a checklist or rating system to fully incorporate the many different ways faculty members contribute to the education of our students and the success of our college. Thus, for each of the 5 dimensions, we described multiple examples of activities and behaviors that meet, exceeds or does not meet expectations. The final version of the dimensions and descriptions for each level of performance is provided in the Appendix.

Seeking Faculty Feedback

The next step in the development of these new faculty evaluation dimensions was to introduce the framework to the business faculty for feedback. We first sent the document to the faculty via email. At the following faculty meeting, we briefly discussed the task force objectives and the process used to develop the performance dimensions and descriptions. Although a few questions arose at the meeting, the task force wanted to create additional opportunities for faculty to express their questions and concerns.

The task force invited faculty to informally discuss the proposed evaluation framework over food and beverages. We decided to meet in small groups to encourage an honest and thoughtful exchange of ideas, so we offered multiple meeting times. No administrators attended these meetings. Sixty percent of the faculty attended one of the meetings we offered. The meetings led to lively discussions

about the task force proposal and the faculty evaluation process in general. The teaching dimensions and descriptions dominated the meetings; no changes were made to the task forces' original recommendations regarding service. With respect to the teaching dimensions, some of the suggestions were minor, such as those involving specific wording or adding items to the dimension descriptions. Other concerns reflected more fundamental differences about what constitutes effective teaching and service. Significant time was spent debating the validity of SETs.

A couple of participants were content with the school's heavy reliance on SETs. These faculty expressed a belief that, on average, students could judge effective teaching. Others disagreed, suggesting that students were not trained to evaluate teaching and were susceptible to a variety of biases. There is research that supports both viewpoints (e.g., Clayson, 2013; Wright & Jenkins-Guarnieri, 2012). Several faculty members noted that the new standards would be a success if they replace the current emphasis on student evaluation ratings with a more balanced view of teacher effectiveness.

If SET scores are used to make personnel decisions, the question becomes, what ratings define "meets expectations" and "exceed expectations"? At Millsaps College, the teacher ratings on most items average around 6 on a 7-point scale. On the rating form, a 6 is described as "excellent" for overall ratings of the course and instructor or "typically accurate" for such behavioral items as "is well prepared". Some participants in the meetings argued that a rating of 4 or above should meet expectations because, if the evaluation form is taken literally, this rating reflects "average" on overall quality or "moderately accurate" on desirable teacher behaviors. Ultimately, the task force decided to include SET scores in the dimension "Quality of Instruction". We recommended that SET scores could be used to help evaluate either "meets expectations" or "does not meet expectations". Under the standard "meets expectations," we wrote "[the instructor] received satisfactory student ratings (for example, as measured by a rating of 5 or above, or within a standard deviation of the college mean)". We added that all of the 23 items on the student evaluation form needed to be considered, as opposed to relying solely on an overall quality rating. This decision reflected a compromise we hoped all sides could accept. Importantly, the task force recommended that SET scores be considered in addition to numerous other instructor behaviors and activities.

We were surprised when one participant took issue with the dimension, "Innovative Practices". This faculty member suggested that innovation was not relevant for some classes. Most survey participants gave innovative teaching practices a high rating (see Table 1), so the task force believed it was important to keep. It was possible, however, that our description of each level of performance (i.e., does not meet, meets, exceeds expectations) on this dimension was deficient. For example, another faculty member complained that we had defined innovation too narrowly with a focus on technology. We modified the description of this dimension, including phrases such as "experimented with new . . . teaching strategies," and "used realistic and current examples, cases, simulations and other exercises," which should be applicable to all instructors, including those who teach highly technical, content-intense courses.

A couple of faculty members commented that the descriptions for each teaching dimension were biased in favor of experiential learning and community engaged learning. A few people noted that lecturing is a valuable teaching method, especially for content that is difficult to understand. The task force acknowledged this and added the phrase, "[the instructor] was especially adept at helping students understand difficult concepts" as an example of "exceeds expectations" under the dimension, "Quality of Instruction".

The final major concern expressed involved the validity of the dimensions and descriptions for each level of performance. A faculty member asked whether we could statistically validate the instrument that we use for evaluation. It is important to note that this framework is not a rating instrument or a checklist. It provides a list of behaviors and activities that help administrators and faculty understand what is expected. Nevertheless, it could be argued that the framework has content validity. It is consistent with the evaluation items listed in the literature and is aligned with our own faculty's expert opinions about what constitutes effective teaching and service. Because this framework clarifies the types of behaviors and activities that are expected for each level of performance, it has to be more valid than our current system, which offers only the vague requirement for "sustained, noteworthy teaching" and "serious attention to the duties and responsibilities of a faculty member," as written in the faculty handbook.

Perhaps the most important consideration is whether this framework will help administrators and faculty better understand expectations for faculty performance and provide them with a tool for identifying areas of strengths and concerns. In fact, this framework has already begun to engage our faculty in discussions about what constitutes effective teaching and service. For example, the conversations about classroom innovation, the use of lecture and community engagement encouraged critical thinking about pedagogy and the connection between classroom experiences and the school's mission and strategic plans.

Decision-makers must make sure they apply this framework in a manner that increases perceptions of procedural, interactive and distributive justice because doing so leads to better work outcomes (Erdogan, 2002; Flint, 1999; Holbrook, 2002) and reduces the chances of illegal discrimination. In addition, administrators should regularly evaluate ratings, promotion and pay decisions to ensure that no adverse impact has occurred. Although the Supreme Court ruled that formal statistical validation is not a requirement for cases of discrimination in personnel decisions, courts will review a variety of evidence to determine whether the decisions were both fair and valid (Lee, Havighurst & Rassel, 2004).

Based on the feedback from the small group meetings, the task force revised the proposed framework and sent the revision to the faculty via email for final review. At the following faculty meeting, those present unanimously accepted the dimensions and descriptions, with the exception of one abstention. The framework will now be used to guide faculty as they write their annual reports and the dean as she evaluates those reports. The final version of the framework is in the Appendix.

CONCLUSION

Faculty evaluation is mandated by various accreditation bodies such as SACS and AACSB, and if done effectively, can provide clear expectations for performance and assist with faculty development. Currently, an overreliance on SETs and vague standards of performance introduce considerable bias and uncertainty into the evaluation process, promoting dissatisfaction among faculty. We have described a process in which faculty were directly involved in the refinement of a business school's faculty evaluation standards. Although the task force did not eliminate the subjectivity associated with faculty evaluations, the framework that was developed makes clear the behaviors and activities associated with not meeting, meeting and exceeding expectations for faculty performance in teaching and service. Because we relied heavily on faculty input from the beginning of the development process, faculty satisfaction of the evaluation process should increase. The almost unanimous vote to adopt the framework is one indication of faculty satisfaction. The steps we took and the framework we developed for evaluating faculty offer direction for others engaging in review and refinement of the faculty evaluation process. The final product will vary by institution, depending upon a school or department's mission, strategic plan and learning environment. Nevertheless, many of the issues and concerns addressed in this paper are universal and the process we used to clarify our standards may also bear good results for others.

REFERENCES

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G., & Spooner, F. (2004). Student Evaluation of College Teaching. College Teaching, 52(4), 134-141.

Association to Advance Collegiate Schools of Business. (2013). Eligibility Procedures and

Accreditation Standards for Business Accreditation. Retrieved from http://www.aacsb.edu/~/media/AACSB/Docs/Accreditation/Standards/2013-business-standards.ashx.

Chonko, L. B. (2006). An Essay on Wisdom in the Teaching Evaluation Process. Marketing Education Review, 16(30), 1-13.

Cipriano, B., & Riccardi. R. (2005). An Analysis of CSU Faculty and Chairs Perceptions of Faculty Performance. Academic Leader, 21(12), 3-8.

Clayson, D. E. (2009). Student Evaluations of Teaching: Are they related to what Students Learn? A Meta-analysis and Review of the Literature. Journal of Marketing Education, 31(1), 1-15.

Clayson, D. E. (2013). Initial Impressions and the Student Evaluation of Teaching. Journal of Education for Business, 88(1), 26-35.

Engelland, B. T. (2004). Making Effective use of Student Evaluations to Improve Teaching Performance. Journal for Advancement of Marketing Education, 5(Winter), 40-46.

Gabris, G. T., & Ihrke, D. M. (2001). Does Performance Appraisal Contribute to Heightened Levels of Employee Burnout? Public Personnel Management, 30(2), 157- 172.

Holbrook Jr., R. L. (2002). Contact Points and Flash Points: Conceptualizing the Use of Justice Mechanisms in the Performance Appraisal Interview. Human Resource Management Review, 12(1), 101-123.

Honeycutt, E. D., Thelen, S. T., & Ford, J. B. (2010). Evaluating and Motivating Faculty Performance: Challenges for Marketing Chairs. Marketing Education Review, 20(3), 203-214.

Kluger, A. N., & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, A Meta-Analysis, and a Preliminary Feedback Intervention Theory. Psychological Bulletin, 119(2), 254-284.

Lang, J. W. B., & Kersting, M. (2007). Regular Feedback from Student Ratings of Instruction: Do College

Teachers Improve Their Ratings in the Long Run? Instructional Science, 35 (3), 187-205.

Lee, J. A., Havighurst, L. C., & Rassel, G. (2004). Factors Related to Court References to Performance Appraisal Fairness and Validity. Public Personnel Management, 33(1), 61-77.

Marsh, H. W., & Roche, L. A. (1997). Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility. American Psychologist, 52 (11), 1187-1197.

Miller, J. E. & Seldin, P. (2014). Changing Practices in Faculty Evaluation. Academe, 100(3), 35-38.

Millsaps College Faculty Handbook. (2007) Revised.

Nasser, F. & Fresko, B. (2002). Faculty Views of Student Evaluations of College Teaching. Assessment & Evaluation in Higher Education, 27(2), 187-198.

Peterson, R. L., Berenson, M. L., Misra, R. B., & Radosevich, D. J. (2008). An Evaluation of Factors Regarding Students' Assessment of Faculty in a Business School. Decision Sciences Journal of Innovative Education, 6(2), 375-402.

Rynes, S. L., Gerhart, B., & Parks, L. (2005). Personnel Psychology: Performance Evaluation and Pay for Performance. Annual Review of Psychology, 56(1), 571-600.

Shao, L. P., Anderson, L. P. & and Newsome, M. (2007). Evaluating Teaching Effectiveness: Where We are and Where We Should Be. Assessment & Evaluation in Higher Education, 32(3), 355-371.

Shepherd, C. D., Carley, S. C., & Stuart, R. S. (2009). An Exploratory Investigation of the Periodic Evaluation Processes for Marketing Faculty: A Comparison of Doctoral-Granting and Non-Doctoral-Granting Universities. Journal of Marketing Education, 31 (2), 143-153.

Shields, P. O. (1996). The Citizenship of Marketing Faculty: An Investigation of Service Activities. Marketing Education Review, 6(2), 83-89.

Silva, K. & Thomsen, D. (2013). Faculty Assessment and Evaluation: Fair, Formative, and Focused. The Department Chair, (Summer), 9-11.

Simpson, P. M. & Siguaw, J. A. (2000). Student Evaluations of Teaching: An Exploratory Study of the Faculty Response. Journal of Marketing Education, 22(3), 199-213.

Southern Association of Colleges and Schools Commission on College. (2011). The Principles of Accreditation: Foundations for Quality Enhancement, 5th ed. Retrieved from http://www.sacscoc.org/pdf/2012PrinciplesOfAcreditation.pdf.

Stapleton, R., John, R., & Murkison. G. (2001). Optimizing the Fairness of Student Evaluations: A Study of Correlations Between Instructor Excellence, Study Production, Learning Production, and Expected Grades. Journal of Management Education, 25(3), 269-291.

Stark-Wroblewski, K., Ahlering, R. F., & Brill, F. M. (2007). Toward a More Comprehensive Approach to Evaluating Teaching Effectiveness: Supplementing Student Evaluations of Teaching with Pre-Post Learning Measures. Assessment & Evaluation in Higher Education, 32(4), 403-415.

Wachtel, H. K. (1998). Student Evaluation of College Teaching Effectiveness: A Brief Review Assessment & Evaluation in Higher Education, 23(9), 191-211.

Williams, K. F. & Rhodes. M. T. (2011, October). Chief Academic Officers' Perceptions about Faculty Evaluation. Paper presented at the 83rd Annual Meeting of the American Educational Research Association, New Orleans, LA.

Wright, S. L. & Jenkins-Guarnieri, M. A. (2012). Student Evaluations of Teaching: Combining the Meta-Analyses and Demonstrating Further Evidence for Effective Use. Assessment & Evaluation in Higher Education, 37(6), 683-699.

APPENDIX

Dimensions and Criteria for Assessing Teaching and Service Performance

This document is intended to build a common understanding of what is expected of faculty in the areas of teaching and service. The examples provided for each level of performance reflect the fact that there are a variety of ways to meet or exceed performance expectations. It is impossible to create a checklist or rating system to fully incorporate the many different ways faculty members contribute to the education of our students and the success of our College. Examples provided for each level of performance are therefore not intended to exhaustive; faculty or the Dean of the Else School may identify additional behaviors that exemplify "meets," "exceeds," or "does not meet" expectations. Greater clarity about performance expectations should enhance the communication among faculty and with the dean and is intended to provide a positive resource for faculty development.

Quality of instruction

Meets expectations: Demonstrated competence in teaching learning objectives.

Examples to consider include: had learning objectives that were consistent with the learning outcomes designated for each course taught; met with class regularly; used class meeting time effectively; employed pedagogy appropriate for learning objectives; received satisfactory student ratings (for example, as measured by a rating of 5 or above, or within a standard deviation of the college mean; consider all questions posed on the evaluation instrument); issued fair grades that reflected learning objectives achieved; assigned papers or projects as required for comprehensive exams; participated in the comprehensive exam process; provided students with syllabi that contained learning objectives, grading criteria, a schedule of reading and written assignments; was accessible to students outside of class to clarify instructions, course content or feedback.

Exceeds expectations: Demonstrated excellence in teaching learning objectives.

In addition to items listed under 'meets expectations,' the faculty member demonstrates excellence through examples such as: utilized relevant assignments (case studies, business problems, simulations, community engagement activities) that challenged students to rise to higher levels of performance than typically expected; developed and utilized learning activities that led students to solve actual problems in the community or participate in competitions; was especially adept at helping students understand difficult concepts; utilized a variety of teaching methods, media and/or technologies in ways that highly motivated students to achieve course learning objectives.

Does not meet expectations: Did not effectively teach learning objectives.

Examples to consider include: learning objectives for the course were not clear or consistent with learning outcomes designated for the course; was not prepared for class or did not use class time effectively; there was no distinction made in the content and performance expectations for a class taught to undergraduates and graduates; used learning activities that were not effective in meeting course objectives or did not engage students actively in learning; received unsatisfactory student ratings (for example, as measured by a rating below 5 or below the standard deviation of the college mean; consider all survey questions); issued grades that did not reflect student performance or were inconsistent with grading criteria described in the syllabus; project or paper assigned for the comprehensive exam did not sufficiently assess learning outcomes assigned for the course; did not participate in the comprehensive exam process as needed; syllabus was incomplete, unclear or not followed; was inaccessible to students outside of class to clarify instructions, course content or feedback.

Innovative practices

Meets expectations: Utilized engaging learning activities to meet course objectives.

Examples to consider include: utilized media and technology in an effective way to engage students in problem solving or critical thinking; taught an established course abroad; used realistic and current examples, cases, simulations and other exercises to actively involve students in learning the course objectives.

Exceeds expectations: Developed new teaching strategies or utilized creative and engaging learning activities to meet course learning objectives.

Examples to consider include: developed or utilized learning activities or assignments that encouraged students to identify and/or solve real problems in an organization within the community; experimented with new technology or teaching strategies; taught a new, innovative course; taught a new course abroad; created and implemented new cases, simulations and other exercises to actively involve students in learning the course objectives; significantly revised a course to incorporate new cases, simulations, exercises, media and/or technologies; taught interdisciplinary courses with colleagues from other divisions or disciplines or served as a guest lecturer in classes or events across campus.

Does not meet expectations: Demonstrated no attempt to update content or utilize new approaches to engage students to meet learning objectives.

Examples to consider include: used outdated texts or reading materials; relied heavily on one or two types of learning activities that encouraged passive learning; learning activities did not engage the students in higher level thinking.

Effort or time commitment

Meets expectations: Had teaching commitments that required time and effort consistent with most other faculty in the business unit.

Examples to consider include: taught an average number of preparations; taught classes that were average in size (15-18 students in this case); supervised a small number (one to three in the case of Millsaps College) directed studies/internships, supervised an honors program student or a student in a program intended for undergraduate students interested in college teaching to shadow a faculty member; provided quality advising/mentoring to assigned students; agreed to teach in summer school or the college's Executive program when asked; taught at night or less desirable times slots; submitted information about student progress in a timely manner for assessment.

Exceeds expectations: Had teaching commitments that required more time and effort than what is typically required for faculty in the business unit.

Examples to consider include: taught more preparations than the average number required; taught more students per class than the average faculty member; supervised numerous directed studies/internships and perhaps also an honors program or shadowing student; taught one or more courses he/she had never taught before; taught in several graduate level courses; taught in the college core curriculum; taught in several less desirable times slots, such as multiple night classes in a semester; attended seminars, colloquies, symposiums, professional meetings, etc., to improve teaching effectiveness; filled in for a colleague who was unexpectedly absent for a significant period; shared teaching materials and/or techniques with other colleagues; guided students in community engagement learning activities.

Does not meet expectations: Spent minimal effort on teaching duties in ways that led to an increased burden for colleagues and/or had a negative impact on the quality of students' experiences.

Examples to consider include: refused to teach required or elective courses in the fall or spring semesters as needed, or refused to teach in the times needed; refused to teach new courses as needed; cancelled classes frequently without justification; was inaccessible to students for advising.

Contribution and leadership

Meets expectations: Demonstrated serious attention to the duties and responsibilities of a faculty member.

Examples to consider include: regularly attended all committees assigned and engaged in the work of the committee; submitted assignments or action items on time; provided quality advising and career support to students; supported efforts to satisfy accreditation requirements; served as advisor for student organizations; participated on temporary task forces or project teams such as search committees; participated in off-campus initiatives such as the unit's community outreach program; may have organized, lead, or otherwise participated significantly in local field trips and/or out-of–town student trips; served the community outside one's official role as faculty member; provided consulting services in a manner that brought attention to the business unit; or served as a reviewer for academic journals; supported assessment efforts.

Exceeds expectations: Initiated projects or programs and/or provided leadership on committees or as the director of a program that led to significant improvements and the enhanced reputation of the business unit, the college and/or the profession.

In addition to items listed under 'meets expectations,' the faculty member demonstrates excellence through examples such as: played a key role (as a committee chair or member, a program director, or an individual) in an initiative that enhanced the quality of the student's educational experience, financial security of the business unit or college, or reputation of the business unit or the college; advised students in off-campus competitions; led temporary task forces or project teams such as search committees with valuable results; played an instrumental role in community outreach efforts such as the unit's community engagement program; engaged in outreach to alumni that led to job opportunities for students or enhanced fundraising efforts; served as an officer or chair for professional organizations; served as an editor for an academic journal.

Does not meet expectations: Played no role in helping the business unit or the college achieve their goals or advance their programs and/or reputation.

Examples to consider include: did not regularly attend or share the workload of the committees to which s/he was assigned or elected; failed to submit or was consistently late in submitting assignments or action items related to committee work, assessment, accreditation, etc.; was not accessible to advisees and did not respond to their emails; did not attend major college and unit events, such as graduation, awards ceremonies, important unit events, etc.; did not complete tasks that were assigned by the Dean or other superior.

Team player

Meets expectations: Was communicative, cooperative and respectful to colleagues on committees, in the unit and across campus.

Examples to consider include: willingly participated on committees as needed; listened to colleagues with an open mind; compromised when appropriate; demonstrated respect for colleagues; responded to requests for data in a timely manner; attended major college and unit events, such as graduation, awards ceremonies, important unit events, etc.; supported on-campus recruitment events.

Exceeds expectations: Words and actions created a more collegial environment on committees, in the business unit and across campus; behaviors enhanced communication, problem solving and commitment to committee, divisional and/or college goals.

Examples to consider include: volunteered to fill roles or committee assignments; actions and initiatives brought people together from across campus; diffused conflict when it occurred in meetings or among colleagues; stepped in for colleagues on committees or class when they were unable to fulfill their role; in addition to major college and unit events, attended numerous faculty, student, prospective student, and/or alumni gatherings, receptions, sporting events, etc.; participated in off-campus recruitment events.

Does not meet expectations: Words and actions damaged collegiality and were detrimental to progress on committee, business unit and/or college goals and problem solving.

Examples to consider include: refused to take on roles and tasks when doing so would have helped spread the workload among colleagues; after agreeing to do tasks, did not follow through; words and actions fueled conflict within committees and among campus groups; refused to compromise when appropriate; dominated discussion in meetings and refused to consider other ideas or perspectives; rarely attended major college and business unit events.