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 utmost 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-
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 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 37% percent 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.P.A. 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 and objectives for the annual evaluation process. The task force was envisioned as a function of the communication between the supervisor and the employee during the performance appraisal 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. Interpersonal and distributive justice issues are largely up to the administrators who would use the evaluations, 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 (Hollander, 2005).

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 and just by all members.

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 Gabel & Inkele (2001), performance appraisals are often used to 1) provide feedback on teaching and service; 2) influence tenure and service promotion; 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 standards for promotion, tenure and merit pay decisions. Our intention was to create a tool that would help faculty clearly understand performance expectations.

Both streams are useful in identifying the factors 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 expectations.

One 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, Ryves, Gerhar & 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. Ryves et al. also indicated that interrater reliabilities are low for behavior-oriented evaluation and the link between pay and performance is generally perceived as weak when such assessment is used. Validity can be the result of poorly constructed evaluation instruments, rating 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 are 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, interpersonal justice and distributive justice are strongly associated with employee affective reaction to performance appraisal, including responses related to morale, satisfaction, commitment to the organization and intentions to leave (Erdogon, 2002; Flinn, 1999; Holbrook, 2002). Perceptions of procedural justice increase when employees are aware of the standards of performance, the standards are consistently applied and the employee has input into the process. Interpersonal 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.

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 used to 1) provide feedback on teaching and service; 2) influence tenure and service promotion; and 3) make merit 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 and improve teaching effectiveness (Clayson, 2004; Hors- ecuyt, Thelen and Ford 2010; Stark-Wroblewski, Ahler- ing, and Brill 2007; and Williams and Rhodes 2002). In this 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 Ras- dowsich 2008; Chonko 2008; Algozzine, Beattie, Bray, Flowers, Greer, Howley, Mohanty, and Spooner 2004; Engelland 2004; Nassar and Fresko 2002; Stapleton and Thorpe 2005). Performance appraisals are often used to 1) provide feedback on teaching and service; 2) influence tenure and service promotion; and 3) make merit decisions. The task force recognized that these three objectives were relevant for faculty evaluation. 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. Interpersonal and distributive justice issues are largely up to the administrators who would use the evaluations, 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 (Hollander, 2005).
Three other studies are helpful in identifying the factors that college administrators typically use to evaluate teaching performance. Honeycutt et al. (2010), asked market- ing chairs from AACSB-accredited institutions to assign 100 points among five factors to evaluate faculty teaching performance. Responses were categorized based on institutional importance given to teaching and research. Among institutions that focused on teaching or had a balanced focus on teaching and research, the following activities were identified: (1) service to school/college, (2) service to department, (3) service to university, (4) service to discipline, and (5) service to community. 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 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 survey by Zhao, Anderson, and Newcombe (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 important. 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: (1) academic advising, (2) 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’ self-evaluations, classroom observation, review of syllabus 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 had a balanced focus on teaching and research, the following activities were identified: (1) service to school/college, (2) service to department, (3) service to university, (4) service to discipline, and (5) service to 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 importance: (1) service on college-wide committee; (2) academic advising; (3) service on departmental committee; (4) departmental administrative duties; and (5) chair 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 50% 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 Riccar di’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 societies as important, whereas 76% of chairs cited it as a 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 is 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 college and/or university, and service on departmental committees. In fact, 67% 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 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 the 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 considered important for evaluation of 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 factors: 4 assessing teaching, 11 advising, and 11 assessing service. Faculty were asked 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 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 administrative support is provided 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,” and “Teaching a New Course.”

Faculty & Chair Input

Our task force generated a list of items that were considered important for evaluation of 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 factors: 4 assessing teaching, 11 advising, and 11 assessing service. Faculty were asked 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 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 administrative support is provided 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,” and “Teaching a New Course.”

Identifying Performance Dimensions

Relying heavily on the survey results, the task force proceeded with development of a multi-dimensional evaluation framework for teaching and service performance. In reviewing those teaching items that received more than 10 (75%) faculty endorsements as important for evaluating teaching performance, it became apparent that a first dimension identified, Quality of Instruction, is indicated by the faculty’s belief that both
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 Elsie 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. The meetings led to lively discussions.

The Student Ratings of Teaching Behaviors and Student Overall Rating of Instructor should carry weight in assessing 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 Revisited 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; Englendall 2004; Wachter 1998; and March 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 1: Evaluating Teaching and Service Performance

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

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; Englendall 2004; Wachter 1998; and March 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.
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 force’s original recommendations regarding ser-
vice. 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. Others described substantial, fundamental differences about what constitutes effective teaching and service. Sig-
ificant 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 teach-
ing. Others disagreed, suggesting that students were not trained to evaluate teaching and were susceptible to a va-
riety of biases. There is research that supports both view-
points (e.g., Clayson, 2013; Wright, Jenkins-Guarnieri, & Neely, 2012). Several faculty members noted that the new stan-
dards would be a success if they replace the current em-
phasis 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 expecta-
tions” and “exceeds expectations”? At Millsaps College, the teacher ratings on most items average around 6 on a 7-point scale. A rating of 6 is described as “ex-
cellent” for overall ratings of the course and instructor or “typically accurate” for such behavioral items as “is well pre-
pared” (e.g., Clayson, 2013). Faculty members 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 desir-
able instruction. Some feel that a rating of 3 reflects “needs improve-
ment,” whereas a rating of 2 is described as “unsatisfactor-
y” (e.g., Clayson, 2013). The task force de-
cluded 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 expec-
tations,” 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 course 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 mem-
ber 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 be-
lieved 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 com-
plained 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 ...” and “used current examples, cases, simulations and other exercises,” which should be applicable to all instructors, including those who teach highly technical, content-intensive courses.

A couple of faculty members commented that the descrip-
tions for each teaching dimension were biased toward the framework 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 under-
stand. The task force acknowledged this and added the phrase, “[the instructor] was especially adept at helping students understand difficult concepts” as an example of “meets expectations” under the dimension, “Quality of Instruction.”

The final major concern expressed involved the validity of the dimensions and descriptions for each level of perfor-
ance. A faculty member asked whether we could statisti-
cally validate the instrument that we use for evaluation. It is important to note that the instrument is not a rating instru-
ment or a checklist. It provides a list of behaviors and activities that help administrators and faculty understand what is expected of them. It is 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 con-
stitutes effective teaching. The task force de-
cluded 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 expec-
tations,” 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 course 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 mem-
ber 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 be-
lieved 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 com-
plained 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 ...” and “used current examples, cases, simulations and other exercises,” which should be applicable to all instructors, including those who teach highly technical, content-intensive courses.

A couple of faculty members commented that the descrip-
tions for each teaching dimension were biased toward the framework 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 “meets 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 the instrument 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 of them. It is 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. The task force de-
cluded 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 expec-
tations,” 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 course 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 mem-
ber 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 be-
lieved 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 com-
plained 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 ...” and “used current examples, cases, simulations and other exercises,” which should be applicable to all instructors, including those who teach highly technical, content-intensive courses.

A couple of faculty members commented that the descrip-
tions for each teaching dimension were biased toward the framework 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 “meets 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 the instrument 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 of them. It is 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. The task force de-
cluded 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 expec-
tations,” 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 course 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 mem-
ber 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 be-
lieved 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 com-
plained 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 ...” and “used current examples, cases, simulations and other exercises,” which should be applicable to all instructors, including those who teach highly technical, content-intensive courses.

A couple of faculty members commented that the descrip-
tions for each teaching dimension were biased toward the framework 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 “meets 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 the instrument 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 of them. It is 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. The task force de-
cluded 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 expec-
tations,” 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 course 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 mem-
ber 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 be-
lieved 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 com-
plained 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 ...” and “used current examples, cases, simulations and other exercises,” which should be applicable to all instructors, including those who teach highly technical, content-intensive courses.
Developing a Multi-Dimensional Evaluation Framework for Faculty Teaching and Service Performance

Diane F. Baker, Walter P. Neely, Penelope J. Prenshaw, & Patrick A. Taylor

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


Southern Association of Colleges and Schools Commis- sion on College. (2011). The Principles of Accredita-


APPENDIX

Dimensions and Criteria for Assessing Teaching and Service Performance

This document is intended to build a common under- standing 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 vari- ety 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 mem- bers 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 ad- ditional behaviors that exemplify “meets,” “exceeds,” or “does not meet” expectations. Greater clarity about per- formance expectations should enhance the communica- tion 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 objec- tives that were consistent with the learning out- comes designated for each course taught; met with class regularly; used class meeting time effectively; employed pedagogy appropriate for learning objec- tives; received satisfactory student ratings (for exam- ple, 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 compre- hensive 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 in- structions, course content or feedback.

Exceeds expectations: Demonstrated excellence in teaching learning objectives. In addition to items listed under ‘meets expecta- tions,’ the faculty member demonstrates excellence through examples such as: utilized relevant assign- ments (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 learn- ing 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 learn- ing outcomes designated for the course; was not pre- pared 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 un- dergraduates and graduates; used learning activities that were not effective in meeting course objectives or did not engage students actively in learning; re- ceived unsatisfactory student ratings (for example, as measured by a rating below 5 or below the stan- dard deviation of the college mean; consider all sur-

vey questions); issued grades that did not reflect stud- ent performance or were inconsistent with grading criteria described in the syllabus; project or paper assigned for the comprehensive exam did not suf- ficiently 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 ac- tivities to meet course objectives. Examples to consider include: utilized media and technology in innovative ways to encourage students to solve new problems and critical thinking; taught an estab- lished course abroad; used realistic and current examples, cases, simulations and other exercises to actively involve students in learning the course objec- tives.

Exceeds expectations: Developed new teaching strat- egies or utilized creative and engaging learning ac- tivities 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, exer- cises, media and/or technologies; taught interdisci- plinary 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 at- tempt 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 learn- ing; 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/interimships; 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/interimships 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.

Meet 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 in 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, led, 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; the unit, the college and/or the profession.

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; the unit, the college and/or the profession.

Examples to consider include: voluntarily participated on committees and 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.

Examples to consider include: rarely attended major college and business unit events.