Community College Strategies

Student Assessment of Teaching as a Component of the Overall Assessment Process:
The Teacher Evaluation Assessment Survey

Terri M. Manning, Denise H. Wells, Lynn Ahlgrim-Delzell

Assessment of teaching effectiveness is an important component of a comprehensive effort to assess institutional effectiveness and student learning outcomes. A critical facet of assessing teaching is student evaluation of instruction. Most "homegrown" teaching evaluation instruments, however, are created or used with little consideration for their role in the overall assessment process. This article describes a case in which the design and implementation of student evaluation of the instruction process have been incorporated in the wider college assessment model.

Central Piedmont Community College began revising its process for assessing teaching in 1999. Because community colleges are primarily teaching institutions, it is obviously important that they have a tool to effectively assess teaching. Our teaching assessment tool was designed to provide faculty with feedback from students that can be used to improve teaching, to give chairs and division directors consistent indicators of student perceptions of faculty (especially part-timers), to provide one of several means of assessing teaching for merit pay, and to contribute to the comprehensive college assessment effort.

A committee of faculty and support staff reviewed the existing instrument for student evaluation of instruction, searched the literature for existing instruments with good psychometric properties, and consulted stakeholders at the college. After performing this initial research, the committee decided to revise the existing twenty-item instrument. The old tool had been judged inadequate; it contained many questions about classroom management (office hours, use of syllabus, and so on) and few questions about teaching effectiveness. The committee found that most teaching evaluation tools used by institutions of higher education were created internally, did not always start at the beginning (answering the essential question "What is good teaching?") and were not tested for validity, reliability, or sensitivity.

The committee evaluated over three hundred items from approximately fifteen teacher evaluation surveys from both community colleges and universities. Once items were selected, revised, or developed, the committee conducted a brief pilot study involving focus groups in approximately ten classes. They revised the tool, experimented with two different scoring rubrics, and conducted a large-scale pilot test before the tool was fully implemented at the college.

We now examine the psychometric properties of the revised Teacher Evaluation Assessment Survey (TEAS) to provide information on the adequacy of the new instrument to evaluate instructors' teaching effectiveness.

The TEAS contains thirty-six items, broadly divided into three sections relating to instructor, course, and student. Each item is rated on a 5-point Likert scale: 5 = excellent/exceptional, 4 = above average, 3 = average/adequate, 2 = below average, 1 = very poor/inferior. The committee, wishing to examine the various aspects of teaching, created items based on five subscales: The Art of Teaching, The Science of Teaching, The Business of Teaching, The Course, and The Student. The five subscales of the TEAS have the following descriptions and point scales:

1. The Art of Teaching involves the more innate aspects of teaching that are not considered methodological. Examples include a teacher's ability to motivate students, be enthusiastic, have a positive attitude toward students and the course, encourage participation, and make students feel valued and comfortable when they ask questions (11 items, possible score range 11–55).

2. The Science of Teaching involves methods or areas that can be taught, such as organizing class time, clarifying materials with examples, making relevant assignments, using the textbook, and teaching new things to students (5 items, range 5–25).

3. The Business of Teaching involves items and issues required by the institution, such as handing out syllabi, applying policies and being fair to students, meeting with the class for the entire period, holding office hours, providing feedback, and announcing tests in advance (6 items, range 6–30).

4. The Course has to do with the characteristics of the course, its applicability to the students' field of study, and difficulty level (3 items, range 3–15).

5. The Student is about the amount of effort and preparation the student puts into the course. While faculty are not responsible for this, it may help to explain the variance in teacher evaluations (6 items, range 6–30).
Psychometric Properties
The psychometric results presented here are based on a sample of 17,041 students across approximately 1,550 course sections, seventy departments, and seven deans using the TEAS in Fall 2000.

Validity. Face validity and content validity were established by using the judgment of experts in the field and by comparing survey results to historical data (using previous assessments and subjective supervisor input). A factor analysis was also completed, using the principal axis method and correlation matrix with a varimax rotation. Initially, the analysis was forced into five factors to test the theory behind the development of the five subscales of the TEAS. The TEAS revealed two strong factors and one weak factor that explain 64.5 percent of the variance in responses to the items. Because the three teacher subscales (The Art of Teaching, The Science of Teaching, and The Business of Teaching) were highly related, the factor analysis separated the items into three major categories (instructor, course, and student) rather than the five subscales. It stands to reason that if a teacher does not use good methodology or keep accurate records, motivate students, or give adequate feedback, it is highly unlikely that students will consider them good teachers. Eigenvalues by factor were as follows: Factor 1 (Instructor) eigenvalue = 19.35; Factor 2 (Course) eigenvalue = 2.61; and Factor 3 (Student) eigenvalue = 1.26.

Reliability. Internal consistency was assessed to determine how consistently the instrument assessed teaching quality across the items. Cronbach’s alpha produced a coefficient of .974, indicating very high internal consistency. Spearman-Brown (split-halves) coefficient was calculated to determine how well the halves correlated with each other. The TEAS yielded a Spearman-Brown (equal-lengths) coefficient of .907, also indicating very high internal consistency.

Sensitivity. Because single-item Likert scales of 1–5 produce ratings that leave little room for improvement, establishing the sensitivity of the TEAS was an important issue. Is a faculty member with a mean of 4.73 really a better teacher than a faculty member with a mean of 4.66 on a given item on a teaching evaluation tool? To be sensitive, a teacher evaluation tool must be able to spread responses among all of the Likert-scale choices in order to discern small differences. In addition, students have to be able to discern small differences between instructors and within one instructor over time. Individual items on the TEAS produced skewness at the top end of the answer scale, with 80 percent of students giving their instructors “4” or “5” on most items. To improve sensitivity, subscales were created, thus increasing the range, variation, and spread of scores. Once means for each subscale were established, skewness at the top end was still observed, but more variation in scores existed (Table 1).

When faculty members demonstrated movement over time within a subscale, those differences could be detected. Instead of seeing change occur across thirty-six separate 5-point scales, instructors could see change across five varying point scales.

Issues Inherent in the Assessment of Teaching
Three critical issues in regard to the assessment of teaching using a particular instrument are (1) the definition of “good teaching” as measured by the instrument, (2) the ability to differentiate between poor and good teachers, and (3) the use of results.

Good teaching isn’t the same for all teachers, nor is it the same for all students. Good teaching is something often defined when it is not there or absent. Creating an assessment scale that doesn’t penalize instructors with different teaching styles is critically important. Faculty were heavily involved in the development of the TEAS, and more attention was given to valued teaching principles (for example, motivating students, encouraging interaction) than to teaching style (for example, lecture, hands-on interaction).

Institutions need to determine whether they want a teaching assessment tool to identify those faculty who need most improvement in their teaching so that appropriate interventions can be designed and offered. This process would allow institutions to establish an acceptable level of teaching quality for all faculty. It is easier for a tool to pick out teachers in the bottom 20 percent of the scale than it is to differentiate between levels of good and excellent teachers. The tool described here adequately identifies those instructors in need of mentoring. Using the subscales, it is also possible to discern changes in performance on the part of good and excellent teachers.

Once a tool is developed, the appropriate use of results is critical in helping managers motivate faculty and aid them (continued on page 13)

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Subscale</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Art of Teaching</td>
<td>48.9</td>
<td>8.1</td>
<td>11</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>The Science of Teaching</td>
<td>22.2</td>
<td>3.5</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>The Business of Teaching</td>
<td>26.8</td>
<td>3.9</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>The Course</td>
<td>12.8</td>
<td>2.4</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>The Student</td>
<td>26.2</td>
<td>2.3</td>
<td>6</td>
<td>30</td>
</tr>
</tbody>
</table>
in improving their teaching. The question of whether standards for determining teaching effectiveness should be developed at the institution, college, department, or course level is not a question that can be answered with statistics. One problem that often occurs when setting standards is that department chairs or division directors often want to set norms for an assessment tool by using the department mean and standard deviation to set the benchmark for good teaching. This process creates a moving target for faculty attempting to make improvements in their teaching by continually raising the benchmark as the department mean increases. Because mean scores can vary greatly from department to department, two teachers from different departments with exactly the same score might be considered above average in one department and below average in the other. Instead, institutions should set a good teaching criterion by determining the acceptable level as measured by their teaching evaluation tool. This benchmark should not change, thus giving faculty an established target measure. While results on the TEAS will vary by department and course, the expected quality of instruction should be an established standard that applies across an institution.

Terri M. Manning is associate vice president for institutional research and Denise H. Wells is senior research analyst at Central Piedmont Community College. Lynn Ahlgrim-Delzell is project director, department of special education at the University of North Carolina—Charlotte.