UNIVERSITY LEADERSHIP COUNCIL

Defining and Measuring Course Quality through Student Course Evaluations

Custom Research Brief • *February 24, 2011*

RESEARCH ASSOCIATES
David Hamstra
Ehui Nyatepe-Coo

RESEARCH MANAGERS
Josh Albert
Sarah Moore

TABLE OF CONTENTS

THE ADVISORY BOARD COMPANY
WASHINGTON, D.C.

I. RESEARCH METHODOLOGY

Project Challenge

A member institution approached the Council with the following questions:

- *How do institutions define a quality course and successful teaching?*
- *What assessment tools do institutions use for student course evaluations? How are such assessment tools designed to provide quantitatively robust feedback?*
- *What measureable standards do student course evaluations use to assess teaching efficiency and course quality? Are these standards consistent across an institution or do they differ between academic departments?*
- *How do institutions use the feedback gathered from student course evaluations? How heavily is student course evaluation feedback factored in the tenure and promotion process at institutions?*

Sources

- Education Advisory Board's internal and online (www.educationadvisoryboard.com) research libraries
- *The Chronicle of Higher Education* (http://chronicle.com)
- National Center for Education Statistics [NCES] (http://nces.ed.gov/)

Research Parameters:

The Council interviewed administrators who oversee assessment or faculty development at large research universities with successful student course evaluation systems, as identified through secondary research.

A Guide to Institutions Profiled in this Brief

Institution	Location	Approximate Enrollment (Total / Undergraduate)	Classification
University A	Southeast	12,900 / 6,900	Private: Research Universities (very high research activity)
University B	Northeast	10,300 / 4,200	Private: Research Universities (very high research activity)
University C	Southwest	48,700 / 38,800	Public: Research Universities (very high research activity)
University D	Southwest	35,000 / 27,400	Public: Research Universities (high research activity)
University E	Northwest	45,900 / 32,700	Public: Research Universities (very high research activity)
University F	Midwest	41,600 / 29,900	Public: Research Universities (very high research activity)
University G	Great Plains	25,800 / 19,500	Research Universities (very high research activity)

II. EXECUTIVE SUMMARY

Key Observations:

- **Student course assessments are one tool institutions use to evaluate course quality and teaching effectiveness.** Other common elements include peer reviews of classes, portfolios of scholarly work, and faculty student advising load.

- **Course evaluations at contact institutions typically have two components: a quantitative section, in which students record their responses to a series of descriptive statements or questions, as well as a qualitative free response section, in which students provide written feedback on courses and/or instructors.** Quantitative data is often more robust than qualitative data, and is thus typically used more frequently in the faculty tenure and review process.

- **When crafting student evaluations, contacts rely on one of three models: a) a central office (usually the office of institutional research and assessment) creates a standard list of questions, which individual departments or faculty members supplement with customized questions; b) a central office creates unique assessments for each distinct instructional format or academic subject; and c) evaluations include no standard questions but are designed by each academic school or department.**

- **Contact institutions provide several promising strategies for developing and sustaining effective student course evaluations, including suggestions for improving evaluation content, promoting teaching development, ensuring validity of responses, and increasing involvement in course evaluations.**

- **In order to ensure validity of responses, contacts recommend performing reliability analyses on the quantitative elements of student assessments**. This is especially important when data is used in decisions surrounding tenure and/or faculty salaries.

- **Institutions with psychometrically sound evaluations often factor evaluation data more heavily in the tenure and review process, whereas institutions with less statistically rigorous assessments may give evaluations less weight.** Regardless of the validity of the data, contacts stress the importance of viewing student evaluations as one part of a faculty member's overall review.

III. Developing a Student Course Evaluation

Elements of a Student Course Evaluation

Student course evaluations are comprised of two major components: quantitative evaluation and qualitative evaluation. Quantitative sections include a series of questions or statements about the course or instructor, which solicit responses on a numeric scale. Qualitative items are often open-ended, with students providing written feedback on topics such as their least and most favorite aspects of a course. During the review and tenure process, some institutions rely more heavily on quantitative results than on qualitative results, as quantitative data tends to be more consistent and robust.

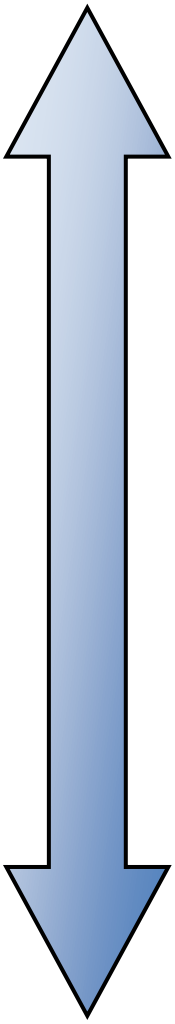
Student Course Evaluations: Overview of Quantitative Components		
Institution	**Quantitative Questions**	**Quantitative Scoring**
University A	Questions cover several topic areas, including course rigor and organization.	Items are scored from 1 to 9
University B	Questions are designed to be discipline-specific, with distinct quantitative items for science/engineering and humanities courses.	Items are scored from 1 to 7
University C	Several questions pertain to the structure of the course.	Items are scored from 1 to 5
University D	Questions address course organization, learning environment, and level of student engagement.	Items are scored from 1 to 4
University E	Evaluations ask questions about overall course rating, course content, and the instructor's teaching effectiveness.	Items are scored on a six-point scale, from "very poor" to "excellent"
University F	Each academic department creates its own quantitative questions.	Items are scored from 1 to 5
University G	The assessment includes items in several areas including learning outcomes, workload, and instructor's grading efficiency.	Items are scored from 1 to 6

III. Developing a Student Course Evaluation

Evaluation Models

Institutions employ one of three models for designing student course evaluations: a) a central office (typically the office of institutional research and assessment) generates standard questions, which are supplemented with customized items provided by individual departments or faculty; b) a central office creates unique forms for each course format or subject; or c) each department or instructor creates a unique student course evaluation.

Standardized



Customized

Model A: Standardized Basic Questions with Customized Additions
University A*, *University C*, *University D*, *University G

At these contact institutions, a central office creates a set of questions that is included on every student course evaluation across the university, with individual departments or faculty members providing additional customized questions. At **University** D, the student course evaluation includes twelve quantitative questions; faculty members can also select additional pre-written qualitative questions, or create their own. Instructors at **University C** are required to include about three to five questions that have been approved by the student government and faculty senate on all student course evaluations; individual departments select remaining questions to also incorporate. A similar process exists at **University A**, where course evaluations consist of a set of core standardized questions and department-specific questions. At **University G** each academic college creates a standard list of questions, which all departments within that college must use. Departments may also supply additional questions.

Model B: Standardized Forms for Distinct Course Formats or Subjects
University E, University B

A central office at contact institutions which follow Model B creates student evaluation forms for each distinct course format (e.g., lecture or seminar), or each academic subject. The office of educational assessment at **University E** offers separate forms for 12 instructional formats: laboratory, distance learning, and large lecture courses, among others. Contacts at **University B** designed two separate student evaluation forms: one for science and engineering courses, and another for humanities, arts, social sciences, and architecture and planning courses. Three entities do not participate in the campus-wide course evaluation system: the school of management, the departments of mechanical engineering, and the department of electrical engineering and computer science each continue to use student course evaluations designed for their department.

Model C: No Standardized Form
University F

At University F, there is no standardized form or questions for student course evaluations; each department or instructor develops a unique assessment. Student evaluations are developed by individual departmental and are administered by faculty members.

IV. Strategies for Successful Course Evaluations

Improving Evaluation Content

…by Defining Course Evaluation Objectives

Evaluation designers should clearly define what they hope to measure through course evaluations when structuring assessments. For example, if evaluations will primarily be used for faculty improvement rather than in the tenure and review process, qualitative questions soliciting descriptive responses might be more applicable. Contacts at the **University D** focus their evaluation on measuring teaching effectiveness rather than overall course effectiveness. In contrast, contacts at the **University F** maintain teaching effectiveness is difficult for students to evaluate and prefer to assess learning outcomes.

…by Limiting the Number of Questions

Including too many questions on a course assessment may cause students to lose interest and rush to finish, rather than give each item careful consideration. Centrally-administered course evaluations at **University B** have roughly 30 questions; however, contacts believe this number may be too high. In order to prevent survey fatigue, the course evaluation at **University B** includes only 12 items. Contacts have also designed alternative questions, which can be substituted for the current items, in hopes of preventing student apathy.

Promoting Teaching Development

…by Offering Resources to Improve Instruction

Ideally, student course evaluations highlight instructors' strengths and areas for improvement. Contacts recommend coupling these evaluations with internal resources for faculty development. **University of F** has different teaching and learning centers for academic fields, including a model for the sciences and social sciences called Teaching as Research, which encourages instructors to evaluate their teaching as they do their research. For example, this might include collecting data about how well students meet learning outcomes. Instructors with low scores on the student assessment at **University D** can go to the center for learning enhancement, assessment, and redesign to improve their teaching effectiveness.

IV. Strategies for Successful Course Evaluations

Ensuring Validity of Responses

Several contacts stress the need for ensuring the validity of quantitative responses, particularly when student course evaluations may impact the faculty tenure and review process. **University D** and **University E** have developed especially rigorous student evaluations:

Course Evaluations at the *University E*

History: Since the 1970s, the office of educational assessment at **University E** has administered the paper-based course evaluation for all undergraduate courses. Assessment staff initially offered five separate forms for different instructional formats, but have since expanded this number to twelve. The evaluation contains quantitative and qualitative questions, though only the quantitative items are scored.

Quantitative Items: Each form includes four standardized items for evaluation: overall course ranking, course content, instructor contribution to the course, and teaching effectiveness. Students rate these items on a six-point scale, from very poor to excellent, and the assessment staff calculate the responses for each standardized item.

Reliability Analysis: Central office personnel conduct periodic reliability analyses on the standardized items. Contacts explain this analysis is used to ensure the items are reliable for individual classes and for comparison across the institution or across courses. In recent years these reliability analyses have been conducted less frequently, as the office has found the results to be fairly consistent.

Reporting Results: Faculty members receive a report from the assessment office for each course, which includes the results of the global items and the qualitative information from open response items. Academic deans receive quarterly reports ranking high and low performing faculty members; they also receive annual reports summarizing evaluations across their academic units by faculty member rank (e.g., lecturer, assistant professor, etc.).

Frequency of Use: The course evaluation is administered annually for 12,000 University E courses. This course evaluation system is also used at over 60 other post-secondary campuses.

Student Evaluations at the *University D*

History: Two years ago, university administrators at the **University D** asked the assessment and measurement specialist to develop a campus-wide teaching evaluation survey; the online student evaluation was thus created. The project took three years to develop: the specialist and his team designed the evaluation in the first year, then tested and validated the evaluation in the succeeding years.

Quantitative Items: The evaluation includes three quantitative topical areas, each with four questions: course organization, learning environment, and student engagement. Undergraduates score items on a scale of one to four. The combined scores for these three factors are then calibrated separately on a 0-1000 scale. Teaching effectiveness is also measured on a 0-1000 scale, with 800 and above considered "highly effective," 400-800 considered "effective," and 0-400 as "somewhat effective;" most instructors score in the highest category of effectiveness.

Reliability Analysis: For the past year, contacts have worked with staff in the statistical research unit to develop a psychometric model for testing the validity of the data. Once the model is perfected, the institutional research and effectiveness department will oversee the scoring of evaluations. Eventually contacts believe validity testing will become unnecessary once the consistency of the data is confirmed.

Reporting Results: The institutional research and effectiveness department oversees administration of the evaluation and distribution of the results. Faculty members receive scores for the three quantitative areas along with their overall teaching effectiveness score; senior administrators receive similar data. Furthermore, a state law requires publicly financed institutions to post information online about their instructors; currently the university only posts teaching effectiveness scores.

Frequency of Use: The university administers the course assessments every semester; contacts report receiving 50,000 student responses last semester and hope to receive closer to 100,000 responses in subsequent semesters as response rates improve. The university recently signed a contract to license the evaluations with a private vendor.

IV. Strategies for Successful Course Evaluations

Increasing Involvement in Course Evaluations

…by Soliciting Faculty and Student Input

Faculty and student involvement is critical for the success of student course evaluations, particularly if they are used in the tenure and review process. During the assessment design phase, contacts at **University D** conducted focus groups with students and academic staff to gather their input about potential questions to include in the assessment. When creating a new evaluation, contacts at **University E** recommend creating a faculty committee in order to determine what information would be most useful for faculty to learn from course evaluations.

…by Emphasizing the Impact of Evaluations

Improving the number of students submitting course evaluations is an ongoing challenge for most institutions. In order to collect robust quantitative data, which is often necessary for the tenure and review process, survey administrators must ensure an adequate sample size. The mechanical engineering department at **University B** has an 80% response rate, which is significantly higher than other departments; instructors in mechanical engineering highlight changes in the syllabus or course readings as a result of student input. Contacts believe emphasizing the weight these evaluations hold creates an incentive for greater student participation.

Differing Response Rates for Paper-Based and Online Evaluations

Contacts report that response rates are higher for paper-based evaluation systems than online systems. **University B** formerly used a paper system, which had a 78% response rate; the new online system has a 60% response rate. **University E** continues to use a paper-based course evaluation system to ensure a high response rate, as contacts stress the need for high response rates if the data is used in the tenure and review process. Online evaluations may also be affected by different environmental factors which can skew the data, whereas in-class evaluations offer a more controlled setting that perhaps offers greater data reliability. However, an online system offers the advantage of easier administration and data collection. **University A** currently uses machine-read fill-in-the-oval sheets for student evaluations, which are analyzed by the office of faculty development and excellence. This process requires extensive staff involvement and contacts are considering moving this process online to reduce administrative effort. One method contacts offer to mitigate potential drawbacks of an online system is to offer online evaluations in class, with students filling the assessment out on laptops.

V. Role of Evaluations in the Tenure and Review Process

Each institution or academic department uses student course evaluations differently in the tenure and review process. Given the plethora of assessments, contacts suggest that some senior administrators may rely too heavily on this data when making decisions about granting promotion or tenure to faculty members. Several contacts caution against this trend, arguing that student course evaluations should be one piece of a holistic faculty review, which also includes peer reviews of courses, a portfolio of scholarly work, and an internal departmental evaluation.

Psychometrically Sound Evaluations

Psychometrically sound evaluations pose the greatest temptation for overemphasis on data among academic leaders. The validity and consistency of the data can overshadow its limited importance, with these assessments allowing for easy comparisons between faculty members within and across departments. According to contacts, the evaluation scores at **University D** were not intended for this type of comparison but rather were intended for faculty members to measure their own individual performance over time. At **University G** many departments use quantitative feedback from course evaluations to measure faculty members against each other and against a departmental average, although contacts discourage this practice. Contacts at **University A** report a similar process. The business school at the **University E** is considering implementing a minimum student course evaluation score for its professors to meet, with a goal of having no professors under this minimum score.

Creating Psychometrically Sound Evaluations at *University D*

Contacts at **University D** describe a five-step process for developing a psychometrically sound student course assessment:

1) Develop the student evaluation with a clear goal of what the assessment will measure
2) Create a psychometric model to develop a scaled score for the evaluations
3) Task a department or unit with administering the evaluation
4) Ensure accurate collection and reporting of data both to faculty and senior administrators
5) Conduct validity studies for several years to ensure consistent data

Less Statistically Rigorous Evaluations

Course evaluations with less statistical rigor--those which have not undergone extensive validity or reliability analyses--typically play a less significant role in the tenure and review process. Contacts at **University C** and **University E** stress the limited, though important, role of student course evaluations at their institutions.

> *"Student course evaluations, even if they were perfect, have limited value."*
>
> *-Council Interview*

PROFESSIONAL SERVICES NOTE

The Advisory Board has worked to ensure the accuracy of the information it provides to its members. This project relies on data obtained from many sources, however, and The Advisory Board cannot guarantee the accuracy of the information or its analysis in all cases. Further, The Advisory Board is not engaged in rendering clinical, legal, accounting, or other professional services. Its projects should not be construed as professional advice on any particular set of facts or circumstances. Members are advised to consult with their staff and senior management, or other appropriate professionals, prior to implementing any changes based on this project. Neither The Advisory Board Company nor its programs are responsible for any claims or losses that may arise from any errors or omissions in their projects, whether caused by the Advisory Board Company or its sources.