Best Practices in Using Aggregate Course Evaluation Data

August 2014

In the following report, Hanover Research examines how institutions of higher education use course evaluation data. The report touches upon issues of administration, reporting, and interpretation, based on research literature and the practices of individual institutions.

HANOVER RESEARCH

TABLE OF CONTENTS

EXECUTIVE SUMMARY AND KEY FINDINGS

INTRODUCTION

Student ratings have long been established as an appropriate means of evaluating teaching in higher education.[1] Although such instruments are to some degree controversial, particularly among faculty, the trend appears to be one of increasing utilization,[2] including opening data access to students for purposes such as selecting courses.[3]

Perhaps the major caveat for the use of student ratings is that they should never be the sole source of information about an instructor's or an institution's effectiveness.[4] The scope of this report, however, is limited to examining how institutions use student ratings of teaching, or course evaluations.[5]

The report is organized into two sections:

- **Section I: Administration of Course Evaluations** – This section reviews issues in the design of course evaluation instruments and in the communication of their results, including the ability of course evaluations to measure instructional effectiveness.

- **Section II: Using Course Evaluation Data** – This section reviews the types of decisions institutions make with course evaluation data, and discusses how they compare results across units. It also reviews institutional policies on course evaluations, as reflected in faculty and student handbooks.

KEY FINDINGS

- **Course evaluations can be reliable indicators of teaching effectiveness**. Although course evaluations may be controversial, particularly among faculty, research shows that well-designed instruments can accurately gauge instructional effectiveness. Instruments need not be long or complex, as examples in this report from Ohio State University and the University of North Texas show, but they should reflect an institution's specific instructional goals.

---

[1] See, e.g., Cashin, W. "Student Ratings of Teaching: Recommendations for Use." The IDEA Center. January 1990. http://ideaedu.org/sites/default/files/Idea_Paper_22.pdf

[2] Glenn, D. "Rating Your Professors: Scholars Test Improved Course Evaluations." *Chronicle of Higher Education*. April 25, 2010. http://chronicle.com/article/Evaluations-That-Make-the-G/65226/

[3] Mueller, B. "Students Push for More Access to Course-Evaluation Data." *Chronicle of Higher Education*. April 28, 2014. http://chronicle.com/article/Students-Push-for-More-Access/146203

[4] See Benton, S. and Cashin, W. "Student Ratings of Teaching: A Summary of Research and Literature." The IDEA Center. 2012. pp. 1-2. http://ideaedu.org/sites/default/files/idea-paper_50.pdf (noting that scholars are "almost universal in recommending the use of multiple sources of data" for faculty evaluation).

[5] These instruments go by various names, including "course evaluation," "student evaluation of teaching," or "student satisfaction survey." Although some experts prefer the term "student rating," for the sake of simplicity this report uses the term "course evaluation" to refer to any such instrument. See e.g., Benton, S. "It's All in the Name: 'Student Ratings' Versus 'Course Evaluations.'" The IDEA Center. April 23, 2012. http://ideaedu.org/ideablog/2012/04/it%E2%80%99s-all-name-%E2%80%9Cstudent-ratings%E2%80%9D-versus-%E2%80%9Ccourse-evaluations%E2%80%9D

- **Traditional course evaluations are largely appropriate for online courses as well**. Research has found that instruments used in face-to-face classes produce similar results when used in online courses. However, as traditional forms may not fully capture the unique aspects of online learning, institutions can tailor evaluations to the online environment in various ways, from the addition to traditional forms of online-specific items to the use of forms specifically designed for online learning, such as the e-SIR.

- **Results of course evaluations are commonly reported at the departmental and institutional levels, as well as for individual faculty**. The IASystem developed by the University of Washington, for instance, provides both formative and summative reports for units such as colleges or departments, as well as reports to individual instructors. Smaller institutions such as Calvin College or Houghton College also publish institutional summaries of course evaluation results, as well as providing departmental summaries to relevant administrators (e.g., department chairs).

- **Some institutions are moving to give students access to course evaluation results**. Driven in part by the rise of third-party rating sites such as RateMyProfessor.com, such initiatives largely seek to aid students as they select courses for future terms. Typically, students only have access to a limited number of survey items (e.g., open-ended responses may not be published).

- **Institutions most commonly use course evaluation results for summative purposes, such as tenure or promotion review**. However, it is not uncommon for results to be used for other purposes, such as formative feedback at the individual or unit level. It may be the case that faculty, and even administrators, find course evaluation results more useful for formative than for summative purposes; a survey of stakeholders can discover such beliefs and contribute to a consensus on how course evaluations should be used.

- **Statistical adjustments may be required to compare course evaluation data across units**. The University of North Texas, for instance, designed a course evaluation system expressly to allow "apples-to-apples" comparisons across the University, which relies on sophisticated statistical methods to control for the influence of department or student major. At Calvin College, the administration has reported the median variation between ratings for different sections in order to assist administrators in properly interpreting results.

- **A faculty handbook should set forth relatively specific guidelines for the use of course evaluations**. Whether evaluations are used summatively or formatively, faculty handbooks or other policies tend to answer questions such as whether evaluations must be administered, who has access to the results, and how the results will be used. Student handbooks, on the other hand, tend simply to encourage students to complete evaluations thoughtfully and thoroughly, if they address the topic at all.

SECTION I: ADMINISTRATION OF COURSE EVALUATIONS

This section considers issues in the administration of course evaluations, specifically the design of course evaluation instruments and the communication of course evaluation results. The first subsection discusses the design of course evaluation instruments, focusing, in turn, on their ability to predict instructional effectiveness and to assess online learning. The second subsection reviews the ways in which course evaluation results can be communicated to various constituencies, including faculty, administrators, and students.

ISSUES IN THE DESIGN OF COURSE EVALUATIONS

PREDICTION OF INSTRUCTIONAL EFFECTIVENESS

As an empirical matter, measures of instructional effectiveness that "commonly appear on evaluations" include the items shown below.[6] These areas of emphasis are also reflected in widely used evaluation instruments such as the Student Evaluations of Educational Quality (SEEQ) or the Individual Development and Educational Assessment (IDEA) evaluation system.[7]

Figure 1.1: Common Components of Course Evaluations

▪ Course **content**, including organization and coverage	▪ Course **difficulty** and workload
▪ Instructor's **communication** skills, including clarity	▪ **Assessment** and grading practices in the course
▪ Quality of student-teacher **interaction** or rapport	▪ **Students' self-rating** of their learning in the course

Source: Gravestock, P. and Gregor-Greenleaf, E. "Student Course Evaluations: Research, Models, Trends."

The literature reflects a "general and long-standing agreement . . . that course evaluation instruments can be . . . reliable tools" for measuring teaching effectiveness, particularly when the instrument has been "carefully constructed and psychometrically tested before use."[8]

---

[6] Gravestock, P. and Gregor-Greenleaf, E. "Student Course Evaluations: Research, Models, Trends." Higher Education Quality Council of Ontario. 2008. pp. 13-14.
http://www.heqco.ca/SiteCollectionDocuments/Student%20Course%20Evaluations.pdf

[7] [1] For the SEEQ, see, e.g., Corbalan, M. et al. "Reduction of the Students' Evaluation of Educational Quality Questionnaire." Proceedings of the 2013 Federated Conference on Computer Science and Information Systems. p. 697. https://fedcsis.org/proceedings/2013/pliks/29.pdf
[2] For the IDEA system, see "Student Ratings of Instruction." IDEA Education. http://ideaedu.org/services/student-ratings

[8] Gravestock, P. and Gregor-Greenleaf, E. "Student Course Evaluations: Research, Models, Trends." Op. cit., p. 28.

In particular, to gauge teaching effectiveness, institutions should define effective teaching, according to their pedagogical and instructional goals, so that course evaluation instruments reflect these goals.[9] At the **Rochester Institute of Technology** (RIT), for instance, a review of teaching evaluation programs at other institutions led to the primary recommendation that "a common definition of 'effective teaching' be discussed and agreed upon among all colleges/departments to establish a baseline for evaluating effective teaching." This institution-wide "baseline definition" can then be adjusted to allow for differences between disciplines. As the RIT report notes, "departments should be afforded flexibility around forming their baseline definition for their fields."[10]

Arriving at a working definition of "effective teaching," however, may pose challenges. Researchers have struggled with measuring the validity of course evaluations in part because of "the non-existence of a single criterion for effective teaching." [11] For instance, scholars debate whether teaching effectiveness is best measured by asking students about specific aspects of teaching or simply soliciting an overall, or "global," rating. Some studies have found a "strong correlation" between global ratings and more specific measures of teaching effectiveness, suggesting that the global rating serves as an accurate proxy for more specific measures, and may even be preferable to a potentially incomplete list of specific teaching behaviors.[12]

> *A "global" rating of instructor effectiveness can serve as an accurate proxy for lists of specific teaching behaviors.*

In practice, institutions may find it effective for course evaluations to include both a global rating and questions about specific teaching behaviors. A recent study of the course evaluation form used at **Ohio State University**, for instance, suggests that evaluations need not pose a lengthy list of teaching behaviors in order to reliably measure instructor effectiveness. The Ohio State form includes nine questions about specific teaching attributes and one global rating question, which are shown in Figure 1.2. Overall, researchers found a "strong association" between these 10 items and instructional effectiveness.[13]

---

[9] Ibid., pp. 29-30.

[10] Canale, A. et al. "Evaluation of Teaching Effectiveness: Benchmark Report & Recommendations." Rochester Institute of Technology. November 13, 2012. p. 5.
http://www.rit.edu/academicaffairs/facultydevelopment/sites/rit.edu.academicaffairs.facultydevelopment/files/docs/Evaluation_of_Teaching_Effectiveness.pdf

[11] Zhao, J. and Gallant, D. "Student Evaluation of Instruction in Higher Education: Exploring Issues of Validity and Reliability." *Assessment & Evaluation in Higher Education*. 37:2. March 2012. p. 228.
http://eric.ed.gov/?id=EJ954999

[12] Gravestock, P. and Gregor-Greenleaf, E. "Student Course Evaluations: Research, Models, Trends." Op. cit., p. 32.

[13] Zhao, J. and Gallant, D. "Student Evaluation of Instruction in Higher Education: Exploring Issues of Validity and Reliability." Op. cit., pp. 231, 233.

Figure 1.2: Student Evaluation of Instruction, Ohio State University*

SPECIFIC TEACHING ATTRIBUTES	OVERALL RATING OF INSTRUCTION
▪ The subject matter of this course was well organized. ▪ This course was intellectually stimulating. ▪ The instructor was genuinely interested in teaching. ▪ The instructor encouraged students to think for themselves. ▪ The instructor was well prepared. ▪ The instructor was genuinely interested in helping students. ▪ I learned a great deal from this instructor. ▪ The instructor created an atmosphere conducive to learning. ▪ The instructor communicated the subject matter clearly.	▪ Overall, I would rate this instructor as…

Source: Zhao, J. and Gallant, D. "Student Evaluation of Instruction in Higher Education."
* For specific teaching attributes, responses are given on a five-point Likert scale rating agreement; the overall rating is also on a five-point scale from "poor" to "excellent."

The **University of North Texas** (UNT) also uses an internally developed instrument, the Student Evaluation of Teaching Effectiveness (SETE), which it developed through an extensive process of research, experimentation, and validity testing. One purpose of the SETE is "to predict teaching effectiveness," i.e., to produce a consistent result from semester to semester.[14]

To design the SETE, a committee of faculty, staff, administrators, and students began with more than 3,000 survey items compiled from course evaluations in use at UNT, published surveys, and surveys used at over 100 U.S. universities, from which it narrowed the list down to 28 items representing three "factors," or clusters of teaching indicators.[15]

To accomplish this, the committee compared items to the literature on teaching effectiveness and used focus groups of students and faculty to learn "whether the faculty and students felt that the statements measured teacher effectiveness." After a field test of these 28 items, the list was narrowed further to a final 12 items, which are shown below. The SETE also asks for a number of "overall opinions," which are used for statistical purposes (i.e., model fit) and do not contribute to the instructor's rating.[16]

---

[14] Carriveau, R. and Herrington, R. "Meeting the Challenges of Developing a Teaching Effectiveness Instrument that Measures Courses Across a Campus on a Common Scale." 2010 IUPUI Assessment Institute. Slides 44, 63. http://www.unt.edu/rss/rich/IUPUI/IUPUI_MeetingTheChallengesOfTeachingEffectiveness_v2.pptx
[15] "Development." University of North Texas. https://sete.unt.edu/development
[16] Ibid.

Figure 1.3: SETE Course Evaluation Instrument, University of North Texas

	Organization and Explanation of Materials	SD	D	A	SA
1	My instructor communicates at a level that I can understand.				
2	My instructor communicates clearly the expectations for learning in this course.				
3	My instructor provides materials that help me understand the subject				
4	My instructor identifies relationships between and among topics				
	Learning Environment				
5	My instructor establishes a climate of respect.				
6	My instructor is available to me on matters pertaining to the course.				
7	My instructor creates an environment of mutual respect.				
8	My instructor creates an atmosphere in which ideas can be exchanged freely.				
	Self-Regulated Learning				
9	My instructor is skillful in guiding me to be more self-directed in my learning.				
10	My instructor encourages me to connect course topics to a wider understanding of the subject.				
11	My instructor arouses my curiosity.				
12	My instructor stimulates my creativity.				
	Overall Opinions				
1	I like this instructor.				
2	I am interested in this subject.				
3	I think the classroom was appropriate for this subject.				
4	I think this was a challenging course.				
5	I would recommend a course taught by this instructor.				

Source: University of North Texas
* SD=Strongly Disagree   D=Disagree   A=Agree   SA=Strongly Agree

ASSESSING TECHNOLOGY USE OR ONLINE COURSES

The present state of research suggests that **course evaluations for online courses need not differ substantially from those used for face-to-face classes.** One recent paper summarizes the conclusion of multiple studies of course evaluations: "[face-to-face] and online courses are more similar than they are different." Accordingly, many of the factors measured by traditional course evaluations would be appropriate for online courses as well, such as the quality of student-instructor interactions or how well the instructor communicates course content.[17]

Studies of specific course evaluation instruments support this finding. Thus, a review of the IDEA system comparing its use in online and traditional courses found only "minor differences" in the outcomes from the two delivery modes, and concluded that the system

---

[17] Berk, R. "Face-to-Face versus Online Course Evaluations: A 'Consumer's Guide' to Seven Strategies." *Journal of Online Learning and Teaching*. 9:1. March 2013. http://jolt.merlot.org/vol9no1/berk_0313.htm

"is useful for both online and traditional courses."[18] Similarly, at UNT the SETE instrument was tested "in order to do a comparison of online versus not-online student responses," and statistical modeling "confirmed the usefulness of the SETE survey items for online courses."[19]

> *Traditional course evaluations can be appropriate for online courses, but they may not fully capture the unique aspects of online learning.*

While traditional course evaluations may produce equally reliable ratings for face-to-face and online courses, the possibility remains that these instruments "will not capture elements that are unique to [online courses]."[20] In other words, traditional course evaluations appear to be an appropriate but incomplete measure of the effectiveness of online courses. Thus, a recent study of online course evaluations at **Northern Arizona University** finds that the traditional course evaluation used by the University "effectively assessed online courses . . . with the exception of peer interaction, which is not measured by the university evaluation." The author recommends that, rather than developing a special online course evaluation, the standard instrument simply be modified to include a measure of peer interaction, which was found to be particularly important for online courses.[21]

Adding survey items to a standard evaluation in this manner suggests one of the major approaches to adapting traditional course evaluations for online courses. Overall, four approaches "seem to have the greatest potential," according to one recent review. These include:[22]

- **Adding items to a traditional course evaluation**: Many instruments, including commercially available ones such as Student Instructional Report II (SIR II), IDEA, or SEEQ, allow for the addition of customizable items to the basic evaluation. Such items could be "specifically designed" to address aspects of online courses, such as the use of technology. Some vendors even suggest these types of items themselves.[23] This may be "the most efficient and cost-effective approach" to addressing online aspects.

---

[18] Benton, S. et al. "An Analysis of IDEA Student Ratings of Instruction in Traditional versus Online Courses: 2002-2008 Data." The IDEA Center. December 21, 2010. p. 2.
http://ideaedu.org/sites/default/files/Technical%20Report15pdf.pdf

[19] "Development." University of North Texas. Op. cit.

[20] Berk, R. "Face-to-Face versus Online Course Evaluations: A 'Consumer's Guide' to Seven Strategies." Op. cit.

[21] Culver, M. "Analyzing the Effectiveness of Using a University Course Evaluation Instrument to Assess Online Course Instruction." Northern Arizona University. 2012. pp. 1, 10.
http://nau.edu/uploadedFiles/Academic/COE/About/Projects/Analyzing%20the%20Effectiveness%20of%20Using%20a%20University%20Course%20Evaluation%20Instrument%20to%20Assess%20On.pdf

[22] Unless otherwise cited, bullet points draw from: Berk, R. "Face-to-Face versus Online Course Evaluations: A 'Consumer's Guide' to Seven Strategies." Op. cit.

[23] See "IDEA Student Ratings of Instruction: Using Additional Questions for Online Courses." The IDEA Center.
http://www.ideaedu.org/sites/default/files/using-additional-questions-online.pdf

- **Revision of a traditional course evaluation**: This approach is "an extension" of the above strategy; if the addition of optional online items is felt to be insufficient, items on the traditional course evaluation can be revised to more extensively address the online aspects of the course.

- **Commercially available options**: Distinct from the use of customizable items on a commercially available instrument, at least two vendors offer instruments "designed expressly for online courses," including the e-SIR, a variation on SIR II, and the University of Washington's IASystem.

- **Published course evaluation instruments**: At least three studies have been published that also put forth instruments "designed expressly for online courses." These tend to cover traditional aspects of teaching as well as items specific to online courses.

Similar options exist for measuring the use of technology in traditional courses. The IDEA system, for instance, includes an item on the use of educational technology,[24] and analysis of IDEA results can show how often instructors use technology and whether such use correlates with student progress. However, in the IDEA system, use of educational technology does not affect an instructor's final scores.[25]

COMMUNICATING COURSE EVALUATION RESULTS

COMMUNICATING AGGREGATE RESULTS

Course evaluation results can be reported in a number of ways and for a number of purposes. For instance, in the IASystem, the **University of Washington**'s proprietary course evaluation system which is used by over 40 other institutions,[26] available reports include:[27]

- **Course Summary Reports**: Individual instructors receive a report for each course at the end of each term, which is largely for formative purposes.

- **Hi-Low Reports**: These reports also serve a formative purpose,[28] but typically are distributed to administrators, to "assist [them] in allocating resources and support." Based on average ratings of select items, the reports list the courses with the highest and lowest scores each term, which allows administrators to identify those

---

[24] "Survey Form – Student Reactions to Instruction and Courses." The IDEA Center. p. 2.
http://ideaedu.org/sites/default/files/Student_Ratings_Diagnostic_Form.pdf

[25] Benton, S. et al. "An Analysis of IDEA Student Ratings of Instruction in Traditional versus Online Courses: 2002-2008 Data." Op. cit., p. 1.

[26] "Course Evaluation." University of Washington. http://www.washington.edu/oea/services/course_eval/index.html

[27] Unless otherwise cited, bullet points draw from: "IASystem Reports." University of Washington.
http://www.washington.edu/oea/services/course_eval/uw_seattle/reports.html

[28] "Using IASystem to Make Decisions." University of Washington.
http://www.washington.edu/oea/services/course_eval/decisionmaking.html

faculty who may need support and those who may be able to provide it (e.g., as a mentor).[29]

- **Annual Reports**: Annual reports present results at the unit level (i.e., college, department), and as such are distributed to deans or chairs. The reports include "average ratings by course level," and are intended to serve a formative purpose at the unit level.[30]

- **Five-Year Summary Reports**: Intended to inform "regular periodic review of academic programs," these reports function as summative ratings for academic units, including colleges and departments.[31] The reports include average ratings by department and break down ratings by course level and faculty rank; ratings are shown for the preceding five years, and include institution-wide ratings as well.

Similar types of multi-level reporting can also be found at smaller institutions using in-house course evaluations. **Calvin College**, for instance, formerly used IOTA Solutions to administer course evaluations, but ended the relationship in 2010-2011. The College now uses a simple, paper form, which is "processed entirely within the Provost's office," using a scanner and software that reads the bubble forms.[32] Results are reported in at least three ways:

- **Instructors**: Individual instructors receive a report for each class section taught, which includes raw numbers and percentages for each rating on each question, and the average rating for each question.[33]

- **Departments**: Department chairs receive summaries of evaluation results for the department, which provide a "departmental context" for interpreting the data.[34]

- **Institution**: The provost's office publishes a report summarizing institution-wide results for select items. These results are presented by the average scores of class sections on each question, divided into quartiles (e.g., 25 percent of sections had an average rating below 3.63 for Question 9).[35]

Calvin College also emphasizes that these reports should be used with caution, particularly when making personnel decisions, encouraging departments to use data from multiple sections and to take into account the natural variation in scores, regardless of the instructor.[36] A 2011 report, for instance, notes that, on the five-point scale used at Calvin,

[29] "About the Instructional Assessment System." University of Washington. http://www.washington.edu/oea/services/course_eval/about.html
[30] "Using IASystem to Make Decisions." University of Washington. Op. cit.
[31] Ibid.
[32] Stob, M. "Fall 2011 Course Evaluations." Calvin College. p. 1. https://www.calvin.edu/admin/provost/teaching/CollegewideSummaryFA11.pdf
[33] Ibid., p. 2.
[34] Ibid., p. 3.
[35] See "Course Evaluations: Interpreting the Results." Calvin College. http://www.calvin.edu/admin/provost/teaching/
[36] Stob, M. "Using Student Course Evaluations in Personnel Decisions." *The Academic Bulletin* (Calvin College). January 13, 2012. p. 1. https://www.calvin.edu/admin/provost/bulletin/2011-12/20120113.pdf

the median variation in scores for an individual instructor teaching two sections of the same course is 0.22, suggesting that "not all differences [in scores] . . . are significant or necessarily indicative of a difference attributable to the instructor."[37]

Other institutions follow similar practices. At **Houghton College**, which uses the IDEA system, instructors receive an individual feedback report at the end of the term, which is also sent to the department chair and academic deans. Instructors' results are benchmarked against all other Houghton faculty and, via the IDEA system, against other faculty nationwide. Reports for units, such as academic departments, may be prepared on request.[38] Houghton's institutional research office also publishes an institutional summary for each semester, which includes comparisons with the IDEA system's national benchmarks.[39]

STUDENT COURSE SELECTION

One trend in the communication of course evaluation results has been to make them available to students when they are choosing courses for future semesters. Although a recent survey of faculty at baccalaureate colleges suggests that most still do not disclose course evaluation data, institutions such as Dartmouth, Yale, and the University of Minnesota have all taken steps to make such data available. To some extent, these systems are envisioned as a "more reliable," institutionally controlled alternative to popular rating websites like RateMyProfessors.com.[40]

> *Some institutions are choosing to publish a limited number of course evaluation items to aid students in their course selections for future terms.*

The **University of Washington** already publishes "selected items" from its rating system, IASystem, in a "course evaluation catalog" that is intended "to assist students in selecting courses."[41] **Temple University** introduced a similar system in 2013. As at Washington, only certain items are made available to students (e.g., open-ended responses are not published), and certain types of courses are excluded – at both institutions, evaluations of teaching assistants are not published. Temple's system also predicates students' access to course evaluation data on their completion of evaluations for past courses, creating an incentive that administrators hope will increase evaluation response rates.[42]

---

[37] Stob, M. "Fall 2011 Course Evaluations." Op. cit., pp. 2-3.
[38] "Overview of Houghton's Course Evaluation Program: The IDEA Center's Student Ratings of Instruction." Houghton College. http://www.houghton.edu/ira/course-evaluation-overview/
[39] See "Institutional Summaries by Semester." Houghton College. http://www.houghton.edu/ira/institutional-research-assessment/course-evaluation-program/
[40] Mueller, B. "Students Push for More Access to Course-Evaluation Data." Op. cit.
[41] "IASystem Reports." University of Washington. Op. cit.
[42] "With Course Evaluation Data Now Available to Students, Feedback Comes Full Circle." Temple University. April 24, 2013. http://news.temple.edu/news/2013-04-24/course-evaluation-data-student-feedback-comes-full-circle

SECTION II: USING COURSE EVALUATION DATA

Broadly speaking, course evaluations can be used for either summative or formative purposes. In the former case, evaluations inform decisions such as whether to grant tenure or promotion; in the latter, evaluations are used to improve instructional practices. Some experts, however, suggest that these two purposes should be kept distinct, and that they may even require the use of different evaluation instruments.[43]

In practice, summative purposes, such as tenure and promotion decisions, may be the "most common administrative use of evaluation data."[44] However, Hanover found significant evidence that institutions use course evaluations for other purposes, including providing formative feedback to instructors and comparing data at the unit level (e.g., departments, colleges). This section discusses some of these varied purposes, as well as examining institutional policies about the use of course evaluations.

COMMON USES OF COURSE EVALUATION DATA

HOW ADMINISTRATORS USE COURSE EVALUATION DATA

Although course evaluation data may be most commonly used to inform tenure and promotion decisions, research suggests that such data can also help administrators to "track changes in teaching skills more generally," such as when the data are "aggregated to determine the teaching quality in a department or program in relation to other programs."[45] A study of how administrators use course evaluation data at a major Canadian university identified three major uses for this data:[46]

- Evaluating individual instructors and monitoring changes in teaching quality over time.
- Evaluating instruction at the level of a department or division.
- Assisting with curriculum planning, such as assigning courses to specific instructors.

As this suggests, course evaluations can be used for various purposes. Thus, it is desirable for institutions to define their objectives in conducting course evaluations. A study from the **University of Alaska Fairbanks** (UAF), for instance, suggests that the development of a "consensus on the intended purpose of course evaluation" at the institution is a prerequisite for considering any changes to the system.[47] Similarly, a literature review

---

[43] Gravestock, P. and Gregor-Greenleaf, E. "Student Course Evaluations: Research, Models, Trends." Op. cit., p. 10.

[44] Ibid., p. 12.

[45] Beran, T. et al. "What's the 'Use' of Student Ratings of Instruction for Administrators? One University's Experience." *The Canadian Journal of Higher Education*. 37:1. 2007. p. 29.
http://ojs.library.ubc.ca/index.php/cjhe/article/download/183545/183490

[46] Ibid., p. 35.

[47] Meyer, F. et al. "Assessment of Electronic Course Evaluation Technology and its Applicability to the University of Alaska Fairbanks." University of Alaska Fairbanks. April 30, 2013. p. 2.
https://www.uaf.edu/files/uafgov/Assessment-of-Electronic-Course-Evaluation-Options-for-UAF_fullReport_final.pdf

conducted by Iowa State University researchers concludes that faculty and administrators should develop "a shared understanding of how student evaluation information is used and its purpose at the institution."[48]

A survey of institutional stakeholders may be an effective way to develop such a consensus, and may reveal discrepancies between how evaluations are used and how they are actually perceived. As noted above, for instance, course evaluations may be most often used for summative purposes, such as tenure review and promotions.[49] Thus, the University of Alaska Fairbanks found, in reviewing its own use of course evaluations, that "'faculty evaluation in the context of promotion and tenure' [is] the main application of course evaluation at UAF."[50]

This is also the case at **Owens Community College** in Ohio, where student teaching evaluations form part of the mandated annual performance review for tenure-track and adjunct instructors. [51] However, a 2011 survey of administrators and faculty at Owens found that both groups were much more likely to find course evaluations important for formative uses than for summative.[52]

> *Summative assessment tends to be the principal use of course evaluations, but faculty and administrators may find them most useful for formative purposes.*

Figure 2.1 shows how Owens faculty and administrators, respectively, ranked the various uses of course evaluations. As can be seen, the survey found that teaching improvement and course improvement were by far the top two most commonly cited important uses for course evaluations. While significant minorities of respondents indicated that evaluations are also important for "performance appraisals," far fewer felt that they are important for either program review or tenure and rank review of individuals.

---

[48] "Student Evaluations of Teaching: Guidelines and Recommendations for Effective Practice." Iowa State University. http://www.celt.iastate.edu/teaching-resources/document-your-teaching/student-evaluation-of-teaching/effective-practice/

[49] Gravestock, P. and Gregor-Greenleaf, E. "Student Course Evaluations: Research, Models, Trends." Op. cit., p. 12.

[50] Meyer, F. et al. "Assessment of Electronic Course Evaluation Technology and its Applicability to the University of Alaska Fairbanks." Op. cit., p. 4.

[51] See Board Policy 3358:11-5-22(B)(2)(c)(iv). "Faculty Evaluation Procedures." Owens Community College. p. 2. https://www.owens.edu/trustees/board_policies/11-5-22.pdf

[52] Rathke, D. and Harmon, J. "Purposes of Student Course Evaluations." Owens Community College. March 2011. p. 11. https://www.owens.edu/ie/purpose-stu-evals.pdf

Figure 2.1: Most Important Uses of Course Evaluations, Owens Community College

RANK	ADMINISTRATORS	FACULTY
1	Instructor feedback for teaching improvement (68.2%)	Instructor feedback for teaching improvement (64.1%)
2	Course improvement (63.6%)	Course improvement (64.1%)
3	Program accreditation (40.9%)	Instructor performance appraisals – adjunct (32.5%)
4	Instructor performance appraisals – adjunct (36.4%)	Institutional accreditation (27.8%)
5	Instructor performance appraisals – full-time (31.8%)	Program accreditation (27.8%)
6	Indirect evidence of student learning outcomes (22.7%)	Indirect evidence of student learning outcomes (24.4%)
7	Institutional accreditation (13.6%)	Instructor performance appraisals – full-time (21.4%)
8	Program review (13.6%)	Program review (20.1%)
9	Tenure/rank review (4.5%)	Tenure/rank review (3.8%)

Source: Owens Community College
Percentage indicates respondents who rated that use as among the three "most important."

To provide formative feedback to instructors, some institutions have introduced mid-term course evaluations, which allow faculty to make adjustments to their teaching mid-course. These instruments often contain more direct, open-ended questions that allow students to provide an "honest assessment" of the instructor's teaching, and university policies ensure that such feedback will not influence summative decisions, such as promotion and tenure.[53]

For whatever purposes they use course evaluations, institutions may need to provide administrators with training and support to correctly interpret the data. Researchers have long suggested that the "minimal facility" that many administrators have in interpreting these instruments poses a "major challenge" for their validity,[54] and recent research suggests that course evaluation data are in fact being misinterpreted and misapplied by administrators, as when major decisions are made based on "differences in means small enough to be within the margin of error."[55] As suggested by the example of Calvin College described in Section I, this can be at least partially addressed by providing administrators with summary data such as the median variations in scores between different sections. The broader problem of comparing course evaluation results across an institution forms the focus of the following subsection.

[53] Medina, B. "As Emphasis on Course Evaluations Grows, Professors Increasingly Seek Midcourse Feedback." *Chronicle of Higher Education*. October 30, 2011. http://chronicle.com/article/As-Emphasis-on-Student/129566
[54] [1] Gravestock, P. and Gregor-Greenleaf, E. "Student Course Evaluations: Research, Models, Trends." Op. cit., p. 35.
[2] Beran, T. et al. "What's the 'Use' of Student Ratings of Instruction for Administrators?" Op. cit., pp. 36-37.
[55] Boysen, G. "The (Mis)Interpretation of Teaching Evaluations by College Faculty and Administrators." *Assessment & Evaluation in Higher Education*. 39:6. 2014.
http://www.tandfonline.com/doi/abs/10.1080/02602938.2013.860950

COMPARING RESULTS ACROSS UNITS

There is clearly value in being able to compare course evaluation results across an institution; as a recent study of practices in the United Kingdom notes, a "standard set of survey questions" can be used to "enable effective benchmarking at course and institutional level."[56]

However, comparisons across units must be made with care. As one senior administrator in the aforementioned UK study notes, "courses should not always be compared like-for-like."[57] At least one study suggests that "comparisons with a college mean should be interpreted cautiously," based on research showing that different disciplines tend to produce different mean scores on course evaluations.[58] Another study, of course evaluations at the **University of Washington**, looked at how "average ratings varied by department," finding that departments with more rigorous grading produced lower evaluation scores.[59]

Differences across units can affect course evaluation results, but various approaches can be taken to control for these differences.

Institutions can take a number of steps to address these concerns. At the University of Washington, for instance, guidelines approved by the faculty senate now recommend that, in using course evaluation data for personnel decisions (e.g., tenure, promotion), "course characteristics" such as the level or the department should be taken into consideration, suggesting that "one way to adjust for course types is by choosing similar courses for normative comparisons."[60] Other institutions have even developed sophisticated statistical models that control for variables such as department or student major; the SETE system at UNT, discussed in more detail below, provides one example.

Some course evaluation systems use a combination of generic and department-specific questions in order to enable both cross-unit benchmarking and discipline-specific findings. This is the model used by nationally prominent course evaluation systems such as the Student Assessment of their Learning Gains (SALG) or the IDEA system, which allows departments to "tailor their evaluation forms to emphasize whichever learning objectives

---

[56] "Effective Course Evaluation: The Future for Quality and Standards in Higher Education." Electric Paper. 2011. p. 6. http://www.lse.ac.uk/newsAndMedia/aroundLSE/archives/2011/effectiveEvaluationReport.pdf

[57] Ibid., p. 10.

[58] "Student Evaluations of Teaching: Guidelines and Recommendations for Effective Practice." Iowa State University. http://www.celt.iastate.edu/teaching-resources/document-your-teaching/student-evaluation-of-teaching/effective-practice/

[59] Glenn, D. "Method of Using Student Evaluations to Assess Professors is Flawed but Fixable, Two Scholars Say." *Chronicle of Higher Education*. May 29, 2007. http://chronicle.com/article/Method-of-Using-Student/122298/

[60] "Recommendations for Use of Student Ratings of Instructors in Merit and Promotion Decisions." University of Washington. Spring 2003. http://www.washington.edu/oea/resources/recommendations.html

are most important in their discipline," while still providing "the ability to compare certain scores with a large, nationally normed database."[61]

Alternatively, some course evaluation systems have been purposely designed to be used across different units. For instance, the Teaching and Learning Quality (TALQ) system, developed at **Indiana University** expressly in order to "successfully predict student learning achievement,"[62] was designed as a "one-size-fits-all" questionnaire. [63] Similarly, the SETE system at UNT was constructed in part to allow the University to make "apples-to-apples comparisons . . . across various academic units."[64]

> *The University of North Texas uses sophisticated statistical methods to control for the influence of department and student major on course evaluation results.*

UNT developed its in-house system principally to provide faculty with formative feedback and to give faculty evaluation committees a teaching assessment tool. However, the University recognized that, "if the SETE were to play a larger role in ongoing UNT campus-wide evaluation standards, then it would need to allow for inter-departmental comparisons."[65]

To this end, the SETE committee developed it as a "population (or site based) normative instrument, where across-departmental influences and student-demographic-influences are minimized as much as possible."[66] Specifically, the statistical method of multi-level analysis of variance (ANOVA) is used to control for the influence of department and student major on course evaluation results, such that, in the SETE model, these factors have "non-significant effects" on the rating for general teaching effectiveness.[67] Scores are scaled across the University, so that any given score "has the same meaning in terms of teaching effectiveness" regardless of department.[68] Scaled scores are further grouped into three ranges that indicate levels of teaching effectiveness: highly effective, effective, and somewhat effective.[69]

---

[61] Glenn, D. "Rating Your Professors: Scholars Test Improved Course Evaluations." Op. cit.

[62] Chadha, R. and Frick, T. "Dependability of College Student Ratings of Teaching and Learning Quality." Paper presented at the Annual Conference of the American Educational Research Association. April 9, 2011. p. 4. https://www.indiana.edu/~tedfrick/aera2011/AERA2011ChadhaFrickTALQpaper_FinalVersion.pdf

[63] Glenn, D. "Rating Your Professors: Scholars Test Improved Course Evaluations." Op. cit.

[64] Ibid.

[65] "Development." University of North Texas. Op. cit.

[66] Ibid.

[67] Carriveau, R. and Herrington, R. "Meeting the Challenges." Op. cit., Slides 52, 55-58.

[68] Ibid., Slide 33.

[69] "Development." University of North Texas. Op. cit.

POLICIES GOVERNING COURSE EVALUATIONS

FACULTY HANDBOOKS

As suggested above, different institutions use course evaluations for varied purposes, such as formative or summative assessment of faculty, and institutional policies reflect this variety. This subsection reviews several different examples of such policies, as presented in faculty handbooks or other documents.

Regardless of the approach taken (e.g., summative versus formative), good practice appears to be for the institution to set forth relatively specific guidelines about course evaluations; as one assessment vendor suggests, "better informing instructors about how [course evaluation] data will be used" can help smooth over the tensions created by evaluation and review processes.[70] Hanover's review of institutional policies found that they typically address questions such as:

- Whether faculty must administer course evaluations, and if so, how often;
- Who has access to course evaluation results (e.g., department chair); and
- How course evaluations results will be used (e.g., tenure review).

Policies may be more or less detailed, depending on the system used and the institution's objectives. **Azusa Pacific University** (APU), for instance, has a detailed faculty evaluation system, which assesses faculty in the three core areas of teaching, service, and scholarship. APU uses the IDEA system to administer course evaluations, and IDEA scores play a large role in the teaching component of this evaluation system.[71] The APU faculty handbook addresses a number of aspects of how IDEA scores are used:

- **Number of courses to evaluate**: During their first three years at the University, faculty must administer IDEA in all of their courses. Beyond this term, faculty are limited in the number of courses for which they may use IDEA, in order to control costs.[72]
- **Administration of course evaluations**: The faculty handbook sets out explicit instructions for how faculty are to administer IDEA, such as when to ask for forms from the central administration and when to administer the evaluations to students. The handbook specifies that IDEA scores are first reported to the faculty member's department head, who in turn will share it with the faculty member.[73]

[70] "Using Student Course Evaluation Data for Good." AEFIS. February 8, 2013. http://www.aefis.com/using-student-course-evaluation-data-for-good/
[71] "Faculty Handbook." Azusa Pacific University. March 2013. pp. 82, 84. http://www.apu.edu/live_data/files/111/faculty_handbook_1314.pdf
[72] Ibid., p. 84.
[73] Ibid., p. 85.

- **Using course evaluation scores**: The handbook describes which scores may be used in the faculty member's evaluation (e.g., adjusted versus unadjusted), but does not specify or suggest minimum acceptable scores. The University also allows faculty to use "adjusted scores that have been compared to other classes within a similar discipline," in order to account for factors beyond the instructor's control, such as differences between departments or majors.[74]

- **Use of alternative measures**: The handbook allows for the use of "other measures of teaching effectiveness" beyond the IDEA ratings,[75] and also provides for situations where IDEA evaluations may be unsuitable (e.g., outside the traditional classroom setting). In such cases, instructors or departments may provide for the use of alternative measures such as classroom observations.[76]

Calvin College also uses course evaluations as part of its summative evaluation of faculty. At Calvin, all faculty are expected to administer evaluations in all courses.[77] As set forth in the faculty handbook, course evaluations must be considered as part of the reappointment process for tenure-track faculty. As part of this process, the department chair compiles a dossier to submit to the academic dean; per the faculty handbook, this dossier must include course evaluation results, including "both the numerical summaries of student evaluations and copies of all student forms that contain comments from the most recent semesters (normally at least the previous two)."[78]

Calvin's handbook also stipulates that course evaluations are sensitive documents and must be handled accordingly. At the end of term for which evaluations are completed, they are viewed by the academic dean, who then shares them with relevant department chairs or program directors, as well as with the individual faculty member. The forms must then be returned to the provost's office, where they are kept in a confidential file for at least five years, after which they may be returned to the faculty member. While on file with the provost, access is restricted to individuals at the department chair/program director level or higher, with the exception of the individual faculty member.[79]

By contrast to APU and Calvin College, **St. Olaf College** presents an example of an institution where course evaluations are largely used for formative purposes. Faculty members are "strongly encouraged" to administer course evaluations, but the College does not proscribe a standard form, such as the IDEA system or an in-house version like Calvin College's. Faculty members do have the support of the institutional research office, which provides assistance with the design, administration, and analysis of course evaluations. Results, however, are reported only to the individual instructor, who may share them with others

---

[74] Ibid.., p. 84.

[75] Ibid.

[76] Ibid., pp. 85-86.

[77] See Stob, M. "Fall 2011 Course Evaluations." Op. cit., p. 1.

[78] "Calvin College Handbook for Teaching Faculty." Calvin College. July 2014. pp. 51-52.
http://www.calvin.edu/admin/provost/handbook/Handbook.pdf

[79] Ibid.., pp. 67-68.

(e.g., department chair) at his or her discretion. Course evaluation results are not considered during tenure or promotion reviews.[80]

Although course evaluations do not inform summative assessments at St. Olaf, they nonetheless play a large official role at the College. For all faculty, the use of course evaluations to improve one's teaching can be adduced as evidence of "continuing one's own development as an instructor," which is one of the criteria considered during tenure and promotion reviews.[81] Tenured faculty, on the other hand, are the one class of faculty who are required to administer course evaluations, which serve as one leg of the mandatory, but strictly formative, post-tenure review system at St. Olaf. This process requires tenured faculty to administer course evaluations to at least two courses each year, the results of which must be shared with the department chair. The form used may be chosen or designed by the faculty member.[82]

> At St. Olaf College, the administration of course evaluations is largely at the discretion of individual instructors.

Despite this largely formative approach to course evaluations, St. Olaf does have a system for seeking student input about faculty teaching effectiveness as part of its tenure review process. This "separate process" is administered by the institutional research office using standard forms,[83] and is overseen by the department chair as part of the broader review process. In particular, it is the chair's responsibility to ensure that a "representative sample" of student reviews is included in the faculty member's dossier.[84]

STUDENT HANDBOOKS

Course evaluation policies appear to be less common in student handbooks than in faculty policies, and tend to be less expansive. The main purpose of these policies tends to be to inform students about how course evaluations are used and to impress upon them the importance of completing evaluations thoroughly and thoughtfully, rather than to establish rules that bind student behavior.

At **Georgia Gwinnett College**, for instance, a public four-year institution, the student handbook simply points to the importance of course evaluations and states that "students are expected to evaluate the course and instructor for each class taken each semester,"

---

[80] "Course Evaluation Policies and Procedures." St. Olaf College. http://wp.stolaf.edu/ir-e/course-evaluation-policies-and-procedures/
[81] "Faculty Manual 2014-2015." St. Olaf College. pp. 31-32. http://wp.stolaf.edu/doc/files/2013/12/FacultyManual.pdf
[82] Ibid., p. 51.
[83] "Course Evaluation Policies and Procedures." St. Olaf College. Op. cit.
[84] "Faculty Manual 2014-2015." St. Olaf College. Op. cit., p. 36.

noting that responses are always anonymous.[85] Similarly, at the **Corcoran College of Art and Design**, a small private institution in Washington, D.C., the student handbook states that students "are asked to complete class and faculty evaluation forms" for each course, and notes that "it is important that students be honest and thoughtful with their evaluations." The policy describes how evaluations are used by institutional administrators and notes that faculty members receive copies of the evaluations; it also specifies that they are anonymous.[86] At the **University of Houston**, a large public institution, the student handbook notes that "[student] input is very valuable" and that "course evaluations are simple to do," and asks students to "please take the few minutes needed to thoughtfully complete" the course evaluation for each class.[87]

---

[85] "2013-2014 Student Handbook." Georgia Gwinnett College. p. 110. www.ggc.edu/student-life/get-involved-on-campus/student-affairs/docs/2013-2014-student-handbook.pdf
[86] "Course Evaluations and Retention of Student Work." Corcoran College of Art and Design. http://www.corcoran.edu/student-handbook-2014/other
[87] "Faculty/Course Evaluations." University of Houston. http://www.uh.edu/dos/studenthandbook/enrichment/enrich_faculty.html

PROJECT EVALUATION FORM

Hanover Research is committed to providing a work product that meets or exceeds partner expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.

http://www.hanoverresearch.com/evaluation/index.php

CAVEAT

The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties which extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every partner. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Partners requiring such services are advised to consult an appropriate professional.