

3-1-2012

## DOC 2012-03 Recommendations for Revision to the Process for Student Evaluation of Teaching

University of Dayton. Faculty Affairs Committee

Follow this and additional works at: [https://ecommons.udayton.edu/senate\\_docs](https://ecommons.udayton.edu/senate_docs)

---

### Recommended Citation

University of Dayton. Faculty Affairs Committee, "DOC 2012-03 Recommendations for Revision to the Process for Student Evaluation of Teaching" (2012). *Senate Documents*. 204.  
[https://ecommons.udayton.edu/senate\\_docs/204](https://ecommons.udayton.edu/senate_docs/204)

This Article is brought to you for free and open access by the Academic Senate at eCommons. It has been accepted for inclusion in Senate Documents by an authorized administrator of eCommons. For more information, please contact [frice1@udayton.edu](mailto:frice1@udayton.edu), [mschlangen1@udayton.edu](mailto:mschlangen1@udayton.edu).

## PROPOSAL TO THE ACADEMIC SENATE

PROPOSAL TITLE: Recommendations for Revision to the Process for Student Evaluation of Teaching

SUBMITTED BY: Faculty Affairs Committee of the Academic Senate

DATE: September 28, 2011, rev. November 9, 2011, rev. December 13, 2011, rev. March 1, 2012

ACTION: Legislative Authority

REFERENCE: Article II.B.1.c. Faculty Academic Policies

SEE ALSO: Faculty Policy & Governance Handbook (FPGH) section IV.H [Faculty Policies/University Policy on Faculty Evaluation/Student Evaluation System].

### I. BACKGROUND & PROCEDURE:

In recent years, the Senate has been engaged in a variety of ways on the issue of student evaluation of instruction, with different concerns taking precedence at different times. The student evaluation instrument and its administration have been the primary items of concern. Most recently, in 2009-10, the Student Academic Policies Committee examined the issue in response to student concerns about procedures for administering the current instrument and uncertainty about how the information provided by students was being used, and its designated ERIC sub-committee (Evaluation, Review, and Innovation Committee) reviewed the current student survey instrument. ECAS and SAPC agreed on the need to revise the current instrument. The current instrument is not sufficiently informed by best practices for the evaluation of teaching, and a single survey instrument is not satisfactory to serve both summative and formative purposes. Recommendations from the SAPC and ERIC led to the official statement that is now to be read aloud whenever the instrument is administered. The 2009-10 Year-End Report of the SAPC indicates that they were unable to resolve the issue of whether “the actual purpose of the teaching evaluation form is ... formative or evaluative,” however, and it was decided that the philosophy behind and purpose for soliciting student feedback on instruction should be determined prior to developing any new instrument.

At the beginning of AY 2010-11, ECAS asked the Faculty Affairs Committee to visit the issue of student evaluation of instruction once again, with the specific charge of identifying the philosophy behind and relevant purposes for student evaluation of instruction and making related recommendations regarding its use at the University. Throughout 2010-11, the FAC discussed the purposes of student evaluation of instruction at length. The committee reviewed a number of scholarly articles and other academic publications on the issue (see appended bibliography), reviewed the Academic Senate document 99-7 (Student Assessment of Instruction), invited

David Wright to speak with the committee, and reflected together on our own experiences—as faculty, administrators, and students—with the issue.

A consensus emerged from this year-long discussion about the general purposes of student evaluation of instruction, and the FAC presented an update to the Senate on March 18, 2011. Taking feedback from the Senate discussion and revisiting the materials and discussion notes from the previous year, a sub-committee of FAC (Sheila Hassell Hughes, chair; Andrea Seielstad, and Caroline Merithew) worked in August and September 2011 to draft this formal report and proposal to the Academic Senate. Additional consultation was sought from University Assessment Committee members Deb Bickford and Don Polzella, Steve Wilhoit (LTC), and David Darrow (former president of Academic Senate). The FAC reviewed, revised, and approved the proposal on September 27, 2011. The proposal was submitted to ECAS on September 28, and discussed at their October 6, 2011 meeting. It was determined that revisions needed to be made to the document to clarify main points. The Faculty Affairs Committee, at their October 25 and November 8, 2011 meetings, discussed the revision. A revised proposal was submitted to ECAS on November 9, and, based on additional feedback, was revised and resubmitted on December 13, 2011. The proposal was then discussed by the Senate on January 20 and February 25, 2012. Based on feedback from the Senate (including feedback solicited from faculty, administrators, and students reported by Senators), the FAC reconvened and revised the original proposal. The major changes to this revision are: 1) The formative process is no longer stipulated as a required one for departments and/or units; and 2) The original recommendation for on-line administration is no longer presented as the “default” option.

## **II. RECOMMENDATIONS:**

### **1. SET for Formal, Evaluative Review Processes:**

#### **A. SET Committee:**

- i. A SET Committee should be established by and report to the Senate and charged to identify and propose to the Senate a new instrument (or other appropriate and effective measure) for Student Evaluation of Teaching that is informed by best practices in SET, is in keeping with the guidelines articulated in this document, and is appropriate for university-wide use.
- ii. The SET Committee should be comprised of members with scholarly expertise in SET and should include representatives from the LTC. Depending upon the range of expertise among committee members, it may also be prudent to draw upon external experts to provide broader perspective and/or to provide feedback on a proposed measure in light of best practices in the field and in keeping with our specific institutional mission, student-learning goals, and needs.
- iii. The SET Committee may also make any specific recommendations regarding the larger process for administering and interpreting SET measures that it

deems appropriate to the primary purpose and guidelines articulated here.

#### B. SET Instrument:

Currently, the survey is the most widely accepted SET instrument in higher education, but another kind of instrument might be proposed, and a different sort of method might be explored in the future. The instrument proposed by the SET Committee should:

- i. be adopted, adapted, or newly developed in order to serve the primary purpose of evaluation for formal review (e.g. tenure and promotion, merit, post-tenure);
- ii. be appropriate for use across the University curriculum;
- iii. solicit students' direct observations regarding specific professional practices and behaviors of the instructor (e.g. holding class regularly, as scheduled; distributing a complete syllabus; returning assignments in a reasonable time-frame; and being regularly available for appointments outside of class); and
- iv. invite student feedback about more general items considered directly relevant to teaching effectiveness which research has shown to produce reliable and valid response data.

#### C. SET Administration:

- i. Either the SET instrument itself or the procedures for its administration and/or interpretation of the data must attempt to control for factors of bias and/or externalities to teaching performance beyond the control of the individual instructor. Given the University's diversity goals, special attention should be given to ways in which social and cultural diversity (in instructor/student demographics and/or in course content or approach) may play a role in student feedback.
- ii. The committee should carefully investigate and consider whether electronic administration of the SET instrument might best serve the primary purpose of evaluation for formal review, and, if this appears so, it may make specific recommendations regarding the resources that would be necessary to support an effective and efficient electronic process.

## 2. SET for Formative Processes:

The SET Committee should develop a set of recommendations and resources, in keeping with best practices, to support faculty engagement with student feedback on teaching for the purpose of formative review.

Formative SET procedures are those designed and used primarily for faculty

professional development. Units, departments, and individual faculty should have the freedom and available resources to develop SET processes that they deem most conducive to the on-going formation of faculty as excellent teachers. This recommendation aims to provide a greater number and wider range of opportunities for students to provide feedback to faculty on teaching (the current Midterm Instructional Diagnosis [MID] process is not widely used) and to provide more information about and options for soliciting relevant and formative feedback from students that is not tied to the formal review of faculty.

### **3. Regular Assessment of SET Processes:**

The new SET process should be formally reviewed and assessed by a Senate committee after three years of implementation. After the initial review, it should be assessed at least every 5 years. More frequent assessment may be mandated if substantial changes are implemented as a result of prior assessment. This will allow the opportunity for on-going response to changing teaching and learning practices at UD as well as evolving best practice in SET more broadly.

These recommendations are consistent with the University’s policies on Evaluating Faculty Teaching for the Purpose of Tenure, Promotion, and Merit (Doc 06-08), as well as Review of Tenured Faculty (Doc 06-11). Additionally, these proposed processes focus on engaging the students in the effort to evaluate faculty and improve teaching fairly and informatively.

## **III. TABLE OF CONTENTS FOR APPENDICES**

Appendix A:	Philosophy & Purpose of Student Evaluation of Teaching.....	5
Appendix B:	Value & Validity of Student Ratings of Teaching.....	7
Appendix C:	Appropriate Uses of Student Evaluation of Teaching.....	9
Appendix D:	On-Line Set Administration .....	10
Appendix E:	Bibliography .....	11

## Appendix A

### **PHILOSOPHY & PURPOSES OF STUDENT EVALUATION OF TEACHING**

There is consensus among researchers that students ought to be invited to provide feedback on instruction because they are the only ones, apart from the instructor, who are present for the full instructional experience in any course. This fact is articulated in our current policy (FGPH IV.H). But by what means and for what purposes should such feedback be gained? Student feedback on instruction may serve multiple and varied functions unrelated to the quality of teaching (such as predicting enrollments or providing a sense of “customer service,” etc.).

Whatever possible functions the formal solicitation of student feedback might serve, however, we should focus our efforts on developing and refining a process that best serves our primary purposes. There are two purposes most widely acknowledged by researchers in the field of evaluation (Green, Calderon and Rider, 1998; Marsh and Roche, 1997; Education, 2010), and we affirm these as the primary purposes for student evaluation of instruction at the University of Dayton:

1. To provide feedback to faculty for their own use in developing professionally as teachers (formative purpose);
2. To provide one source of information, among others, for interpretation by administrators and faculty peers in the processes of faculty review and evaluation (summative purpose).

Currently, the single university-wide means of soliciting student feedback on instruction is the student evaluation of teaching survey. As a means of formative feedback, student survey ratings can be helpful to individual instructors, but they have not been shown in and of themselves to be the most effective tool for faculty development in teaching. This is both because individual faculty may not be sure how best to respond to student ratings and because instructors may judge the data collected by standardized measures irrelevant and/or unfair (Berk, 2008; McKeachie and Svinicki, 2010; Williams and Ceci, 1997).

As a means of information for administrative review of faculty teaching, student evaluations have generally taken the form of institutionally standardized surveys that may allow some room for customization by departments or instructors (Education 2010). The ability to compile and store large amounts of information over time is one primary benefit of surveys. The ability to compare data among units within a single institution and across institutions is frequently cited as a secondary reason for standardized surveys (Education, 2010; UD Faculty Policy and Governance Handbook IV.H.7), but there is little reported evidence of such comparisons, either nationally or at UD.

It is not clear that the current survey instrument and process serves effectively in the formative development of faculty at UD; anecdotal data suggests that faculty—especially untenured faculty—experience them primarily as measures for summative external evaluation. Surveys have certainly served their summative purpose at the University of Dayton, but some level of

serve both well because of its content (standardized questions that do not reflect best practice and do not serve varied contexts and specific needs) and form (a paper form administered to every student, in every class, every semester).

The FAC believes that the two primary purposes for soliciting student evaluations regarding instruction need to be addressed substantially and effectively in any new SET process(es) but concludes that this can best be done by severing the two functions.

## Appendix B

### VALUE & VALIDITY OF STUDENT RATINGS OF TEACHING

Researchers in the field and instructional faculty across disciplines have argued extensively about both which specific criteria are pertinent to effective teaching and which criteria students are able to rate reliably and accurately. There is now widespread, though not universal, agreement on some matters, but there is less consensus on others.

As noted above, the UD Faculty Policy and Governance Handbook affirms that students are uniquely positioned, “based on repeated observations” (IV.H.2), to comment upon the teaching practices and habits of their instructors. This is especially important in determining whether professional standards and/or best practices have been employed by an instructor (e.g. holding regular class meetings, as scheduled; distributing a syllabus with course policies and requirements; clearly indicating expectations for assignments; providing feedback on written assignments; and being regularly available to meet with students outside of class hours). Information on such behaviors is important not only in making summative evaluations for purposes of tenure, promotion, and merit pay, but also in identifying areas of concern (or “red flags”) that may require intervention of some kind. These are also the kinds of student-rated items most widely accepted as valid by both researchers and instructors (Williams and Ceci, 1997). A new SET process should therefore include student evaluations about instructors’ specific professional practices. Having reviewed scholarly and professional literature on SET (including articles reporting on primary research findings, reviews of literature on specific issues, and published commentary and discussion by instructional faculty and administrators), and having discussed the issues at length in light of our own observations and experiences as faculty, administrators, and students, we feel equipped to identify dominant trends in the research and to raise points of contestation.

There is substantial agreement in recent research on SET that students are able to provide reliable ratings using more general criteria for effective teaching - such as clarity, organization, enthusiasm, and rapport (Marsh and Roche, 1997; Kulik, 2001) - as well as in offering holistic judgments of teaching performance (Green, Calderon, Reider, 1998). Indeed, student ratings show much greater systematic reliability than do faculty or administrator observations based on classroom visitation (Marsh and Roche, 1997). However, it is with respect to these general criteria and, especially, to holistic judgments, that faculty - including faculty members of the FAC - are most likely to doubt the validity of student ratings (Williams and Ceci, 1997). Some researchers also question the basic fairness of these evaluations, given the absence of experimental data by which to assess validity (Williams and Ceci, 1997). Others raise substantial questions about bias which have yet to be systematically addressed in the research. Potentially biasing elements such as students’ prior interest in the subject, expected grade, class size, and workload or course difficulty have been widely examined and found statistically irrelevant to student ratings by a number of researchers (Marsh and Roche, 1997; Cashin, 1988); nevertheless, there is general consensus that instructors believe student ratings to be tied to course difficulty and grading standards and therefore feel pressure to “dumb down” their courses in response to SET (Cashin, 1988; Williams and Ceci, 1997; Zucker, 2010).



Further, with respect to demographic characteristics of the instructor, research indicates a complex correlation between such factors and other behavior that factors into the student evaluation process. Thus, while there may be some bias resulting from direct demographic characteristics of faculty and students, a more complex set of factors affects the evaluation process. For example, research has established a powerful link between student ratings and certain nonverbal expressions, mannerisms, voice quality and behaviors such as the way a professor smiles, gestures and walks into the room. These mannerisms often stem from physiology, culture, and habit and, as such, are related to race, gender and other identity characteristics. Social stereotypes and perceptions, in turn, filter perceptions of these cues such that students may perceive them differently depending on the professor's race, gender, ability, and other characteristics (Merritt, 2007). This combination of factors injects bias into the assessment of good teaching. Research has shown that these biases and other extraneous factors do make a difference in the student evaluation of teaching process (Johnston, Hammon, and Watson). Significantly, however, research also establishes that the evaluation processes and instruments we use contribute to unreliable and biased results more than do the students who participate. Conventional written assessments conducted in a moment at the end of the semester, for example, may reflect instinctual biases that more contemplative and/or dialogical forms of feedback may avoid.

It is the position of the committee that faculty members should be evaluated based on their teaching effectiveness and that any process or instrument designed for student assessment be controlled for extraneous factors, with special attention paid to the possibility of bias based on instructor identity characteristics such as race, gender, age, and native language. A variety of standardized assessment tools available for use at an institutional level have been assessed for bias and their ability to handle potential biases (Suffolk County Community College). Other methods, such as mid-semester small group focus groups have also been evaluated for their biases and ability to correct for them. As the University presses to advance its diversity goals with regard to recruitment, retention, and instruction, avoiding unfair bias in the review of faculty teaching must be a core concern.

Finally, there is general agreement among researchers that students should not be asked to rate items for which they lack the expertise to make valid judgments. While some items are readily acknowledged as invalid areas for student assessment—such as an instructor's knowledge of the field - there is greater debate about whether students' own learning is something they are equipped to rate (Weinberg, Fleisher, and Hashimoto, 2007). It may be worth asking students whether they believe they learned much in a course, but this should not be considered, in itself, conclusive evidence of student learning.

## Appendix C

### **APPROPRIATE USES OF STUDENT EVALUATION OF TEACHING**

We affirm the current University policy that student-generated data only ever be used as one measure, among multiple measures, in the process of evaluating faculty teaching. This is consistent with our current university documents Evaluating Faculty Teaching for the Purpose of Tenure, Promotion, and Merit (Doc 06-08), as well Review of Tenured Faculty (Doc 06-11). Further, we strongly caution that student feedback be treated not as an evaluative product in itself, but rather as data requiring interpretation by those assigned to make formal evaluations of faculty teaching (Cashin, 1988). Teaching is not a mechanistic project but rather a complex art practiced in the context of human relationships. The skill and impact of teachers cannot ultimately be reduced to numerical measures, and we caution both administrators and committees of peer evaluators against the convenience of treating SET numbers as objective and conclusive data for comparing, ranking, and rewarding faculty performance.

Further, we are concerned that, in using a single, multi-purpose survey instrument, the two primary purposes (formative and summative) for soliciting student feedback are too often intertwined and confused in the minds of students, faculty, and administrators. Given UD's current instrument, which attempts to serve both purposes in a series of questions that are mixed together and, in some instances, not well designed to serve either purpose (e.g. question 9, which asks whether the quantity of work assigned--as opposed to difficulty or complexity of work--was appropriate to course level), we are also concerned that SET responses and their interpretation may be based on irrelevant, misconstrued, or too-limited information.

In addition, those who find the instrument unhelpful to their own goals may neglect the use of formalized student feedback for professional development. Finally, research indicates that faculty are most likely to benefit from formative feedback from students when provided during the course of a term or immediately afterwards, along with consultation regarding ways to respond (Kulik, 2001; Berk, 2008-09; McKeachie and Svinicki, 2010). By contrast, data from SET is most relevant to administrative review of teaching when it is taken as part of a broad data set (averaged results from a range of courses over a significant period of time) and is framed by contextual information about the kinds of courses and students taught (AAUP, 2006).

## Appendix D

### ON-LINE SET ADMINISTRATION

Many colleges and universities across the country are turning to on-line administration of SET survey instruments, for a variety of reasons, including:

- reduced processing time;
- reduced paper waste;
- greater ease in storing data over long periods of time (necessary to summative evaluation);
- greater ease in searching, summarizing, and comparing of numerical data for the purposes of administrative review; and
- greater ease in reading, storing, searching, and summarizing student comments.

UD's pilot testing of on-line evaluation via Isidore showed student response rates ranging from 24% to 100%. The Department of Health and Sport Science, which has been using on-line administration of the standard survey for several years and uses Isidore to track whether individual students have responded or not (but not the content of their responses), consistently earns near a 100% response rate by withholding students' grades until they respond. Other participating departments, such as Chemistry (around 87% response), have been able to achieve strong response rates by simply instructing, encouraging, and reminding students to complete the on-line evaluation. Other institutions have developed a range of communication and incentive strategies for boosting response rates (Academic, 2010). Provided all students have the opportunity to contribute to SET (which is easier to ensure using an on-line survey than an in-class one), response rates do not necessarily need to be extraordinarily high in order to be statistically representative. Further, if compiled across a range of courses over multiple years, as we believe any SET data should be, reasonable response rates would supply sufficient information to identify meaningful patterns.

If a standardized survey instrument is to be used for formal evaluation, then on-line administration best serves that purpose. Faculty who may also want to use standardized SET data for formative purposes will have the benefit of quicker feedback, and, should they desire to add customized questions to an instrument, on-line administration makes that quite simple. Despite these different needs and uses, then, both faculty and administrators would benefit from on-line administration of SET.

## Appendix E

### BIBLIOGRAPHY

#### I. PUBLISHED SOURCES CONSULTED:

AAUP (2006). Statement on teaching evaluation. Policy documents and reports, 10th ed. Washington, D.C. 139.

Berk, R.A. (2008-09). Beyond student ratings: A Whole new world, a new fantastic point of view. *Essays on Teaching Excellence*. 20 (1). Accessed at <http://www.podnetwork.org>.

Cashin, W.E. (1988). Student ratings of teaching: A Summary of the research. Idea Paper 20. Center for faculty evaluation and development. Kansas State University.

Crenshaw, K.W. (1991, July). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43(6). pp. 1241-1299.

Education Advisory Board: University Leadership Council (2010). Gathering and applying meaningful data from student course evaluations. Custom research brief.

Glenn, D. (2010, Dec. 12). Is your psychology 102 course any good? *The Chronicle of Higher Education*. Available at <http://chronicle.com>.

Green, B.P., Calderon, T.G., and Reider, B.P. (1998, Feb.). A Content analysis of teaching evaluation instruments used in accounting departments. *Issues in Accounting Education*. 13(1). pp. 15-30.

Kulik, J.A. (2001). Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research*. 109. pp. 9-25.

Marsh, H.W. and Roche, L.A. (1997, November). Making students' evaluations of teaching effectiveness effective: The Critical issues of validity, bias, and utility. *American Psychologist*. pp. 1187-1197.

McKeachie, W.J. and Svinicki, M. (2010). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers*. Belmont, CA: Wadsworth.

Merritt, D. J. (2007, January). Bias, the brain, and student evaluations of teaching. Ohio State Public Law Working Paper No. 87. Available at SSRN: <http://ssrn.com/abstract=.963196>.

Miller, M. H. (2010, May 6). "Online evaluations show same results, lower response rates." *Chronicle of Higher Education*. Available at <http://chronicle.com>.

Student Academic Policies Committee of the Academic Senate. (2010). Year-End report, 2009-2010. University of Dayton.

Suffolk County Community College Office of Institutional Research. Student ratings of instruction: Report and recommendation. Available at <http://instsrv.sunysuffolk.edu/strate.htm>.

Weinberg, B.A., Fleisher, B.M., and Hashimoto, M. (2007, January). "Evaluating methods for evaluating instruction: The Case of higher education." Cambridge, MA: National Bureau of Economic Research (NBER) working paper No. w12844. Accessed at <http://papers.ssrn.com>.

Williams, W.M. and Ceci, S.J. (1997, September-October). *Change*. pp. 13-234.

Zucker, S. "Evaluation of our courses." (2010, August). *Notices of the American Mathematical Association (AMS)*. 821.

## II. PREVIOUS SENATE PROPOSALS & POLICIES:

Senate Document 00-01. *Demographics: A Report to the Academic Senate from the SAPC*. February 18, 2000.

*Purposes of Faculty Evaluation*. Approved by the Faculty, April 1976; Approved by the President of the University, July 1977. *University of Dayton Faculty Governance Handbook*. p. 62.

Senate Document 98-03. *Revision of Faculty Evaluation Form*. Submitted by the SAPC. December 12, 1989.

Senate Document 99-07. *Student Assessment of Instruction*. Submitted by the SAPC. December 3, 1999. (includes *Rationale for Revising Student Evaluation of Faculty*, and *Student Assessment of Instruction Procedure*)

Senate Document 00-01. "1) *Directions for Completing Student Assessment of Instruction*; 2) *Report of Results of Student Assessment of Instruction*." Submitted by the SAPC. February 18, 2000.

Senate Document I-04-08. *Use of Student Evaluations in Judging Teaching Effectiveness*. Submitted by the FAC. December 3, 2004.