

MEMORANDUM

TO: Fort Wayne Senate

FROM: Mark Masters, Chair
Faculty Affairs Committee

DATE: March 22, 2012

SUBJECT: Student Evaluation Task Force Report

DISPOSITION: To the Presiding Officer for implementation

WHEREAS, the Fort Wayne Senate charged the Student Evaluation Task Force with “formulating ‘a set of standards for IPFW for the procedure and use’ of student evaluations; and

WHEREAS, the Student Evaluation Task Force completed their responsibilities and submitted their report to Faculty Affairs along with creating a web-based resource site (<http://libguides.lib.ipfw.edu/courseevaluations>); and

WHEREAS, Faculty Affairs believes their findings to be well supported and reasonable;

BE IT RESOLVED, That the Fort Wayne Senate endorse the use of the *Recommendations and Best Practices Guidelines*: (<http://lgdata.s3-website-us-east-1.amazonaws.com/docs/1149/357554/StudentEvaluationTaskForceReport.pdf>) from the task force in future use of student evaluations at IPFW.

TO: Mark Masters, Chair of Senate FAC
FROM: Student Evaluation Task Force
DATE: December 16, 2011
SUBJECT: Recommendations from the Student Evaluation Task Force

In spring 2011 the Student Evaluation Task Force was charged with formulating “a set of standards for IPFW for the procedure and use” of student evaluations. The Task Force was to submit its report to the Faculty Affairs Committee of the Senate by December 2011. We have completed this charge and our report is attached for your consideration.

The first part of the charge to the Student Evaluation Task Force was to consider alternatives to the current scanner system. For this charge, the Task Force has sent a memo (attached) to Vice Chancellor McKinney with our recommendations concerning the delivery system as well as an institutional home for the oversight responsibilities for student evaluations on the campus.

The second part of the charge concerned the instrument, procedures for administration of the evaluations, and policy related to how student evaluations are used. In our report we have distinguished between policy recommendations and best practices guidelines. We present two sets of recommendations for the Faculty Affairs Committee to consider implementing as policy: (1) Items Used and Oversight of the Process, and (2) Administration Procedures. Our report also includes a set of Best Practices Guidelines that would not be binding, but should be shared with departments and administrators and sent to the Senate for information only.

We assume that you will send policy recommendations to the Senate as an action item. If you have any questions about the recommendations or guidelines, please don't hesitate to contact Yvonne Zubovic.

Task Force Members:

Elaine Blakemore, COAS & PSY, co-chair
Yvonne Zubovic, MATH, co-chair
Hardin Aasand, ENGL, COAS representative
Tiff Adkins, LIBR representative
Sheena Choi, EDUC, CEPP representative
Cigdem Gurgur, MMK, DSB representative
Rebecca Jensen, NURS, HHS representative
Dina Mansour-Cole, OLS, ETCS representative
Joyanne Outland, MUS, VPA representative
Carol Sternberger, OAA representative

Enclosed Attachments:

Recommendations and Best Practices Guidelines
Memo to Vice Chancellor

IPFW Faculty Senate Student Evaluation Task Force, 2011
Recommendations and Best Practices Guidelines
December 16th, 2011

This document is in three sections. The first section is preliminary remarks about the role of student evaluations at IPFW. The second is a set of recommendations that we are presenting to the Faculty Affairs Committee of the Senate for them to consider implementing as policy that would be binding on faculty and academic departments at IPFW. The third is a set of best practices guidelines that we recommend being made available to departments and administrators, but which would not be binding. There is also an appendix concerning the online administration of student evaluations.

Preliminary Remarks

Over the course of the Spring and Fall semesters of 2011, The Student Evaluation Task Force engaged in a very extensive data gathering process. We examined scholarly literature about student evaluations of teaching, explored systems used by other campuses, and surveyed both chairs and the faculty with respect to their experiences and departmental policies about student evaluations. Results of the data gathering, committee minutes, and other documents can be found on a website constructed for this purpose by Task Force member Tiff Adkins.¹

It is very important to consider the purposes of student evaluations. The two major purposes are often labeled as formative (teaching improvement) and summative (evaluation for purposes of promotion, tenure, and annual review). Based on the survey of faculty, faculty members appear to believe that the formative role of student evaluation data is more important to them (i.e., student feedback helps faculty to improve their teaching and the structure of their courses, and can be used by chairs and other mentors to assist faculty in improving). We also think that is the more important purpose.

With respect to the summative evaluation of teaching, the majority of faculty responding to our survey thought student evaluations ought to be no more nor less important than other measures of teaching effectiveness. However, they also indicated that general practice on this campus is to make student evaluations THE most important measure, and they were not especially satisfied with that state of affairs. That is, the faculty appear to be saying that student evaluations have been overemphasized as a measure of teaching effectiveness for purposes of tenure, promotion, and annual review at IPFW.

For many years IPFW administrators and others have stated that teaching should be evaluated with multiple measures. Experts in faculty evaluation support that position. For example, one widely used resource² suggests 13 ways to evaluate teaching (see the Task Force website for additional references). Student evaluations are surveys of student satisfaction with teaching.

¹ <http://libguides.lib.ipfw.edu/courseevaluations>

² Arreola, R. (2007). *Designing a Comprehensive Faculty Evaluation System* (3rd ed.). San Francisco: Anker Publishing Co.

They are affected by many variables in addition to effectiveness of instruction. It is consistently recommended that they be one measure of teaching effectiveness, but not the only one.

It is also the case that students are capable of evaluating only certain things (e.g., whether information presented is clear, whether instructor is enthusiastic, whether instructor is fair, whether materials are returned in a timely fashion, etc.). They are not capable of determining others (e.g., appropriateness of objectives; instructor's knowledge, its depth, and whether it's current; the instructor's incorporation of department policies or course objectives; the appropriateness of grading standards). Instruments to be completed by students should include only items that students are able to judge, and faculty peers should be the evaluators of items that are deemed important that students cannot judge.³

Student feedback should certainly continue to play a role in the summative review of teaching, but we agree with the majority of the faculty in our survey: student evaluations should be no more nor less important than other measures of teaching effectiveness. Further, our reading of the scholarship on this topic⁴ leads us to conclude that results from student evaluations should generally only be used to make very broad judgments for summative purposes (e.g., exemplary, competent, not competent).

Policy Recommendations

I: Items Used and Oversight of the Process

1. The content of student evaluations should be predominantly determined at the department level. However, having a small number of consistent items is reasonable and desirable, both at the university and college levels.
2. For summative purposes (evaluation for promotion, tenure, reappointment, and annual review) and to ensure comparability across all departments and programs, the university should adopt two core items, generally known as "instructor overall," and "course overall." Sample items from the Purdue Instructor Course Evaluation Service (PICES) are below:
 - a. Overall, I would rate this course as: Excellent - Good - Fair - Poor - Very Poor
 - b. Overall, I would rate this instructor as: Excellent - Good - Fair - Poor - Very Poor.
3. Departments should adopt some items taken from a standardized instrument with known reliability and validity, rather than a locally-developed instrument whose quality is unknown. Note that such instruments often have hundreds of items from which to choose. Examples of such instruments are the PICES⁵ and the IDEA⁶ systems. (PICES items can be used without permission, and incur no cost for their use.)
4. All departments should include some items for summative purposes (evaluation) and some items for formative purposes (teaching improvement).
5. We recommend that the Division of Continuing Studies (DCS) no longer undertake the evaluation of instructors who teach distance learning courses that carry academic credit. Rather, departments should evaluate their own courses and instructors in all cases. DCS may

³ See Seldin, P. (2006). *Evaluating faculty performance: A practical guide to assessing teaching, research, and service*. San Francisco: Anker Publishing Company. (page 56)

⁴ e.g., see McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218-1225.

⁵ <http://www.purdue.edu/cie/web/search/catalog.pdf>

⁶ <http://www.k-state.edu/catl/ratings/idea/index.htm>

certainly wish to survey students taking distance classes as to services they can offer, but the summative and formative review of faculty (full- and part-time) should be done by academic departments.

6. Norm comparisons and comparisons to department means or medians, to other faculty, courses, or departments should be avoided for summative purposes (i.e., for promotion, tenure, and annual review). Rather, departments are asked to develop criteria or standards by which, in their estimation, student evaluation results for their courses reflect competent or exemplary teaching. Under such a set of standards, *in principle*, all faculty might be found to be competent teachers. We note that if this recommendation is adopted it will require a modification to OAA 99-1, which asks for departmental means to be included in P&T cases. In the place of norms or means, the departmental criteria or standards for competent or exemplary teaching would be provided. (Suggestions for the appropriate use of norms are provided below in the “best practices” section.)
7. For summative review of faculty (promotion, tenure, reappointment, annual review), including the evaluation of part-time faculty, student evaluations will typically be part of the data considered. However, they should be only one of several possible measures of teaching effectiveness, and should not be given more weight than other measures.⁷

Policy Recommendations

II: Administration Procedures

1. There should be a standard set of instructions that accompany the evaluations, and they should be printed and included with the packet of evaluations, and read to the class. We recommend these instructions:
“Please use this opportunity to evaluate this course over the entire semester. Your thoughtful answers to these questions will provide helpful information to your instructor and to the department. Please do not talk with other students while you are completing the evaluation. Your responses are anonymous (do not include your name or any identifying information), and will not be provided to the instructor until after final grades have been turned in.”
2. Evaluations should be administered in a regular class period near the end of the semester. For online classes they should be available at some point during the last week or two of the semester.
3. The instructor should **not** be present when the instrument is administered, and the evaluation should be proctored by someone else. The proctor can be a student chosen by the instructor, or a staff member, or someone else chosen by the faculty member or the department. If the proctor is a student in the class, the instructor may read the standard instructions to students prior to leaving the room.
4. We **do not recommend** as a campus policy a process where student evaluations are delivered online for all classes. Many faculty were not in favor of this mode of delivery for face-to-face classes, and the literature on this topic clearly shows a reduction in student response rate that would compromise the validity of the data. Thus, we do recommend that paper delivery of student evaluations continue to be made available. Of course, should particular departments,

⁷ Some may ask what other measures are available for part-time faculty, or for senior tenured faculty who no longer undergo peer review. At the very least, a faculty member can easily report what steps they took to keep their teaching current and/or to make changes in the past year such as incorporating new scholarship in the field in which they are teaching, and/or taking steps to keep current in relevant pedagogical practices.

schools, and colleges choose to deliver all student evaluations online that is their prerogative, but they should seriously consider the threat to the validity of the data that this practice engenders. Because of the importance of this issue, we have attached a summary appendix on this topic.

Best Practices Guidelines

1. We recommend that the university provide resources to maintain normative data (i.e., average ratings—means or medians; standard deviations; ranges) of items over time, at least over a rolling five-year period. (Note: Normative data, especially with respect to items about specific aspects of teaching that can be used for formative purposes—to make decisions about what aspects of teaching to modify—can assist faculty and their mentors in showing areas for improvement. However, it should be noted that small deviations in scores from average ratings should not be overemphasized.)
2. Because departments should be the unit for determining the majority of the content of student evaluations, when practical, we recommend that departments use some common items for all their courses. Online courses, labs, and clinics may follow a different format.
3. We strongly recommend that departments include some items that measure such things as whether the instructor was rigorous or demanding and whether the students learned in the course. Some sample items from PICES⁸ (item number in parentheses) include: “My instructor has high academic standards” (384), “This course supplies me with an effective range of challenges” (402), “My instructor challenges me to think” (412), “I worked harder on this course than on most courses I have taken” (431), “I learned a great deal from my instructor” (433). We found it difficult to find items that seemed appropriate for all courses; we thus thought it better for departments to make their own decisions about what items of this type would be suitable for them.
4. Although we have recommended that departments have their own core of items that are used in all their courses, they should also be encouraged to have flexible instruments that meet a variety of course, instructor, and learning goals, and to support faculty in making flexible choices of items that meet their needs. Instruments such as PICES have hundreds of items from which to choose, and departments and faculty may supplement these with locally-developed items. Having a variety of items across a department’s courses also makes the instrument more interesting for students to complete, and thus they may take it more seriously.
5. Departments may wish to consider including some open-ended items as well as some scaled items on their departmental instrument. Open-ended items often produce more useful information when they ask about specific items (e.g., small groups, lectures, quizzes, the textbook) rather than a general “comments” item. Open-ended student comments are often very helpful for formative purposes; one can get good ideas about how to improve from students’ comments. However, experts in the student evaluation field are very cautious about using student comments for summative purposes. Summative review—review that impacts someone’s job evaluation and future employment—should be based on measures that are well developed, credible, and legally defensible; student comments don’t meet that standard. There is little evidence that they are representative of all students, reliable, or a valid measure

⁸ <http://www.purdue.edu/cie/web/search/catalog.pdf>

of actual teaching effectiveness. Thus we recommend against using student comments for summative review.

6. Department chairs, other administrators, and promotion and tenure committees are urged to be very cautious about using formative items for summative purposes.
7. Chairs, committees, and other evaluators should be mindful of how class attributes impact student ratings. For example, lower level and general education classes are often rated less positively than higher level and graduate classes.
8. Department chairs are encouraged to have a grace period (e.g., a year) before using student ratings to evaluate new faculty, especially those with little prior teaching experience. Of course, this information can certainly be used for formative and mentoring functions by the chair and others. A grace period can also be applied when a more experienced faculty member makes a change to their teaching practices or courses. In all cases, summative judgments are best based on several courses over at least a period of a year, preferably longer.
9. Instruments should not be overly long—an instrument that takes longer than approximately 10-15 minutes to complete is probably too long for most classes. In some cases, however, an instructor may wish to use a longer diagnostic instrument to assist in teaching improvement.
10. Departments should take steps to ensure the security of the data—such that while in transit to the department they cannot be tampered with, modified, lost, or obtained by the instructor prior to completion of class grades.
11. Departments should make every effort to provide timely results to faculty members. Ideally, they should be received no later than two months after the completion of the semester, sooner if possible.
12. Ideally, there should be a proctor-identification-form to be signed by the person administering the instrument indicating the class, date and time, whether the instructor left the room, and whether students completed the form independently and refrained from discussions with other students.
13. Untenured faculty are typically encouraged to evaluate every class, and to make all of these evaluations available to the chair for both formative and summative functions. All faculty are encouraged to evaluate all classes so that students have an opportunity to make their views known. However, departments may want to develop policies such that tenured (or long-time continuing lecturers, clinical faculty, and limited term lecturers) may choose to provide (or have considered) only a subset of these for purposes of annual evaluation.
14. Departments should take steps to make sure that an appropriate number of students complete student evaluations. Various researchers have provided guidelines as to the number of responses needed for the sample of respondents to be representative of the entire class enrollment. Franklin and Theall⁹ indicate that the sample size needed depends on the class size—the smaller the class size, the higher the response rate needed. Their recommendations are provided in the table below. Others recommend that a minimum of 10 students raters and at least 67% of the class are needed for representation.

⁹ Franklin, J. L., and Theall, M. (1991). Communicating ratings results to decision makers: Design for good practice. In M. Theall and J. L. Franklin (Eds.), *Effective practices for improving teaching. New directions for teaching and learning*, no. 48. San Francisco: Jossey Bass

**Student Evaluation Task Force
Recommendations, Page 6**

Class Size	Recommended response rate
5 – 20	At least 80 percent; more recommended
20 – 30	At least 75 percent; more recommended
30 – 50	At least 66 percent; 75 percent or more recommended
50 – 100	At least 60 percent; 75 percent or more recommended
100 or more	More than 50 percent; 75 percent or more recommended

15. Related to this question is class size overall. Sometimes evaluations are not administered in very small classes. Departments must consider and balance four things when judging whether to administer evaluations in very small classes: 1) students having a right to, or at least an expectation of, anonymity; 2) students having an opportunity to provide feedback; 3) faculty members needing to receive student feedback for formative purposes; and 4) faculty members needing data for summative review.
16. Administrators and faculty who make personnel decisions using student evaluation should undergo training as to the appropriate use of this information. We hope that in future, the Office of Academic Affairs or CELT will make such training available on a regular basis.

17.

**Appendix A
Online versus Face-to-Face Administration (for face-to-face classes)**

Many universities are beginning to move to online administration of Student Evaluations of Instruction (SET; Berk, 2006; Pallett, 2006). The table below (adapted from Berk, 2006) summarizes strengths and weaknesses of each:

	Face-to-Face	Online
Cost	More expensive to administer	Higher initial setup cost, then cheaper to administer
Accessibility	Only one available time	More accessible, but should probably have a defined window of availability
Ease of administration		Perhaps easier, but still requires set up by staff
Anonymity	Hand written forms may be less anonymous	All responses are typed
Staff time	More, especially if cleanup and typing are needed	Less, but not zero, responses need to be downloaded and organized
Responses to open-ended items		Tend to be longer, more detailed, more thoughtful
Turnaround time	Slower, especially if open-ended responses are typed	Much faster
Response rate	Typically at least 80%	Can drop much lower unless incentives and encouragement are used. If lower than 65%, not valid measure
Possibility for collaboration	Controlled by proctor	No control over shared completion
Standardization of administration conditions	High	Lower
Ability to modify instrument	More difficult	Simple

Obviously, the main advantages of moving to online SET are lower cost and reduction of staff time to clean and scan forms and type open-ended responses (note that only some departments do this). The main disadvantages are lack of control over the conditions of administration and reductions in the response rates. These reductions can be serious. Pallett (2006) reports a typical drop from more than 80% of students completing forms to 45-55% doing so, a range that would produce high rates of invalid and nonrepresentative results. Nulty's (2008) comparison finds a typical range of 55 to 75% for paper surveys, and 20 to 45% for online surveys, lower in both cases than Pallett's figures, and alarmingly low for online surveys in some cases.

This phenomenon has also been reported in the popular press:

http://www.boston.com/news/local/massachusetts/articles/2010/04/07/colleges_see_decline_in_evaluations_after_going_online/

There are various suggestions for improving response rates, and experimental evidence that they can work (e.g., Dommeyer, Baum, Hanna, & Chapman, 2004). Nulty (2008) lists fifteen recommendations, some of which are included below:

- a. Link the survey in an email sent directly to students (called “pushing” the survey).
- b. Send multiple reminders (especially if they can only go to students who haven’t yet completed it).
- c. Strong encouragement from faculty, with emphasis that responses are important and will be taken seriously. Give directions or advice on how to make constructive criticisms, and include items (usually open-ended) where this can happen.
- d. Provide rewards, small number of points as incentives, access to grades earlier, or drawings for prizes.
- e. Assure students of anonymity.
- f. Make the survey brief.

Note: References can be found on the Task Force Web page at:

<http://libguides.lib.ipfw.edu/courseevaluations>

Task Force Members

Elaine Blakemore, COAS & PSY, co-chair
Yvonne Zubovic, MATH, co-chair
Hardin Aasand, ENGL, COAS representative
Tiff Adkins, LIBR representative
Sheena Choi, EDUC, CEPP representative
Cigdem Gurgur, MMK, DSB representative
Rebecca Jensen, NURS, HHS representative
Dina Mansour-Cole, OLS, ETCS representative
Joyanne Outland, MUS, VPA representative
Carol Sternberger, OAA representative

TO: Vice Chancellor McKinney
FROM: Student Evaluation Task Force
DATE: December 12, 2011
SUBJECT: Institutional Home for Student Evaluations

Part of the charge of the Student Evaluation Task Force was to consider issues related to the delivery and administration of student evaluations. Related to this issue is the question of where the processing center for student evaluations should be housed. In considering the possible “institutional homes” for a student evaluation processing center, the Task Force looked at what other universities, including several of our peer institutions, do. These universities house this function in a variety of offices, all of which report to academic affairs. We also discussed issues related to our specific campus needs and culture. Based on this work, we have the following recommendations:

- (1) The office with oversight for student evaluations should be under Academic Affairs.
- (2) Given the importance of this work, additional staff should be assigned to cover these oversight responsibilities. We recommend not adding these duties to an already heavily burdened staff member.
- (3) We suggest that the oversight of student evaluations be housed within one of the following (not in rank order):
 - a. Assessment Office
 - b. CELT
 - c. Testing Services, but with this office moved from Student Affairs to Academic Affairs

We have investigated several student evaluation systems and evaluated them using criteria related to convenience of use, flexibility of items, availability of norms over time, and paper versus online delivery. Our ranking of these systems, along with an estimated annual cost, is provided in the attached table titled Systems Ranked in Order of Preference. We were reluctant to recommend systems with only online delivery of student evaluations given the persistent problem with low response rates. The literature raises the possibility that incentives to increase these response rates (such as withholding grades) may have limited impact or may result in a negative response by students, calling into question the validity of the evaluations. We recommend that Class Climate be adopted for use, with items drawn from the Purdue PICES library. This system will allow either paper or online delivery of the evaluations, allows maintenance of norms over time, and offers a wide range of items suitable for the needs of the various departments and schools on campus.

Task Force Members

Elaine Blakemore, COAS & PSY, co-chair
Yvonne Zubovic, MATH, co-chair
Hardin Aasand, ENGL, COAS representative
Tiff Adkins, LIBR representative
Sheena Choi, EDUC, CEPP representative

Cigdem Gurgur, MMK, DSB representative
Rebecca Jensen, NURS, HHS representative
Dina Mansour-Cole, OLS, ETCS representative
Joyanne Outland, MUS, VPA representative
Carol Sternberger, OAA representative

Systems Ranked in order of Preference (Final)

	Annual Cost	Convenience and IPFW staff time	Flexibility of items	Availability of Norms over time	Online versus paper
Class Climate (note scanner also needed)	Scanner (but software comes with this system). \$40,000 to \$50,000 per year licensing fees and other startup costs Some staff time here	Some staff time to set up items and print forms, but less than currently. Sounds convenient	High, whatever departments use, or could include some standard items	Yes, and could maintain over time, and use flexibly	Both
Local, online plus paper, staff person to build norms	Scanner and software; paper costs; staff person's salary and benefits	Very convenient to departments; remove a large workload from some departments	High, whatever departments use, or could include some standard items	Yes; that would be one of the main purposes of this person	Both
Status Quo + new software/scanner	Scanner and software cost; Annual cost of paper and staff time in departments	Low, still many hours of local departmental staff time	High, whatever departments use, or could include some standard items	Probably not unless department does it for own use, but not across campus	Both
Local, online, staff person to build norms and oversee	Cost of staff person's salary and benefits	Very convenient to departments; remove a large workload from some departments	High, whatever departments use, or could include some standard items	Yes; that would be one of the main purposes of this person	Online only
Course Eval	\$20,000 the first year; \$16,000 years 2 and 3; no other costs	Removes almost all work from local staff Benefit of access to community of users	High, whatever departments use, or could include some standard items	No, but big enough semester database for decent normative data	Online only
Local, only online	None; very cheap; departmental staff time, hidden cost	Still hours of departmental staff time, but significantly less than above	High, whatever departments use, or could include some standard items	No, unless department does it, but not across campus	Online only
IDEA	Ranges from \$12,000 to \$17,000 annually depending on which form used, and which reports desired	Extremely convenient, very little staff time, basically just ordering and shipping	Need to use IDEA items, short or long form. Could add some items of own, but least flexible of all options.	Excellent maintenance of long-term norms; national database, comparisons to all relevant groups	Both

Scantron forms for evaluations (the ones presently in use) cost about \$50 per 500.