Faculty Senate Working Group on Evaluating Teaching Effectiveness
Report of April 2011

Committee
James F. Young (Professor, ECE and senator) co-chair; Paul Padley (Professor, PHYS) co-chair; David Caprette (Lecturer, BIOS and senator); Karen Nelson (Professor, Jones School); Carolyn Nichol (Lecturer, CHEM); Ann Saterbak (Professor in the Practice, BIOE); and Hadly Wickham (Asst. Professor, STAT).

Charge
- Recommend processes to gather data on teaching and course effectiveness for both formative and summative purposes;
- Recommend strategies and programs that would provide incentives and opportunities for faculty to improve teaching effectiveness;
- Work with the Committee on Teaching to revise its charge appropriately.

Introduction
The membership and the charge of this working group (WG) were not finalized until about the middle of the Fall semester 2010. To some extent this WG represents a continuation of earlier senate working groups on undergraduate teaching, but with a new charge. While we held several meetings, most of our communication was by electronic means, including an OWL-Space collaboration site. Unfortunately, our deliberations never progressed beyond the first item of our charge, how to gather data on teaching and course effectiveness. In late November we divided into two subgroups to examine the issues of student surveys (Young, Caprett, and Nelson) and peer review of teaching (Padley, Nichol, Saterbak, and Wickham). We came together again on April 19, 2011 to discuss and integrate the work of the two groups.

We are unable to provide recommendations for specific implementations or practices for several reasons. We had a diverse set of strongly held opinions on some matters that sometimes prevented complete consensus, and we were unable to get input and information from external sources such as the Student Association, and Rice administration. We are also reluctant to provide strong recommendations because our WG membership does not adequately represent all the schools on campus; we have no representation from humanities or social sciences. Instead, we present here a summary of our work, our findings, and some conclusions that were possible. We hope that these will provide a basis for some initial discussion by the whole senate, and guidance for a new WG to continue this important work next year.
For clarification, in the following sections our usage of the terms *summative* and *formative* assessment align with the educational literature on assessment and evaluation. Summative assessment data is used for decisions regarding instructor official performance reviews, promotions, salary, and other possible consequences. Formative assessment data is used solely by an instructor to improve his or her own teaching effectiveness, and perhaps by the institution to improve course design and structure. Formative assessment data is often made available only to the instructor.

**Student Course-Instructor Survey**

**Problems:**

- According to a study of the present Rice evaluation instrument by Randy Batsell and Rick Stoll, “…all 13 rating items (9 instructor and 4 course) are really only capturing 2 underlying dimensions – one is the instructor and the other is the course. This empirical evidence suggests that we can reduce the number of items being rated by the students and still provide the same quality of information.” Another way of phrasing their conclusion is that the present Rice evaluations provide very little information. Their report points out a number of other issues.

- The present system has such limited reporting or analysis capability that the usefulness of its limited data is further degraded. Extensive research shows that student evaluations have a high level of reliability and validity, but it is also true that, other things being equal, elective courses tend to get higher ratings than required courses, upper-level courses tend to get higher ratings than lower-level courses, small and moderately sized classes get higher ratings than very large classes, and student ratings in engineering and the sciences are lower than ratings in other fields (Felder & Brent, *Chemical Engineering Education*, 42(1), 33-34, 2008). Thus it makes no sense to compare the ratings of a large, required introductory physics class with an elective Shakespeare seminar, or even to the average of all Rice classes. Our present system does not provide useful comparison reports.

**Conclusion:** replace the present Rice course survey.

**Features of a new system**

- Basic comprehensive range of questions that reveals more than two dimensions of effective teaching and course quality, preferably verified by previous studies.

- Some possibility of adding custom questions at times by the instructor, department or school, perhaps just 1–3 questions.
• Wide range of data reporting and analysis functions, far beyond just comparing a course with the average of all Rice courses. The possibility of downloading more detailed data for studies would be an advantage, as would the ability to compare Rice courses with those of peer institutions.

• Ability for the instructor to use or administer the survey anytime during the course, with results going only to her or him, for course improvement.

Possible survey options:


• IDEA, http://www.theideacenter.org/node/5. We note that the IDEA website has many short papers on evaluating teaching, the validity and use of student survey information, etc., each with many references.

• ETS SIR II, http://www.ets.org/sir_ii/about

  We have asked Randy Castiglioni, Associate Vice President for Administration, to provide information on the implementation issues with these systems, including the cost of running the present system. He has not yet responded.

  We have asked the SA to collect data on students’ use of the present survey data, and their opinion on these alternate surveys, including eliminating the posting of open comments (see below). They were unable to collect or provide any information this semester.

Student Open-Ended Comments

Problems:

• The literature on teaching assessment invariably states that student comments on courses and instructors should never be used for summative evaluation (as opposed to numerical survey data), and preferably should be available only to the instructor for formative purposes. In the past, Rice had violated both of these tenets, and continues to make student comments public.

• Many students seem to rely on the open-ended comments to make course selections, viewing them as an adjunct to word of mouth advice from their peers. We do not have good data to support this, but the antidotal evidence is strong. Thus, not posting comments publically would eliminate this source of information for students.

Conclusion: Eliminate the use of student comments in any summative assessment and reward procedures, including teaching awards. Most of our WG believe that student comments should not be public, and, if collected, should go only to the instructor. We note that the Jones school does not make student
comments public. Although our charge does not extend to collecting data to help students pick courses, we are sensitive to their concerns, and do not make such a recommendation at present. We hope, however, that the adoption of a survey instrument that provides a range of information on course and instructor effectiveness, plus other measures, will eventually provide enough information for students to choose courses without the posting of student comments.

**Peer Review of Teaching and Courses**

**Problem:**
A significant body of research indicates that summative evaluation of teaching should be based on multiple types of data. Often a triangulation between student survey data, peer reviews, and a teaching portfolio is recommended. In 2003 the National Academies published an extensive review of the subject (http://tinyurl.com/3kl23yo). While this review was focused on Science, Technology, Engineering and Mathematics (STEM) much of the discussion is broadly applicable. A brief summary of the literature on the topic has been written by Felder and Brent (http://tinyurl.com/39rvpa) and gives extensive references. Eliminating student comments from summative evaluation at Rice means that there will be only one source of data: numerical results of student surveys. This in itself strengthens the need for a new comprehensive student survey with broad reporting capabilities. But that is not sufficient.

**Conclusions:** Rice should (1) institute some type of peer review process for teaching, and (2) permit more extensive teaching portfolios, with a format suitable for the particular school, in promotion, salary, and teaching award procedures. Information and examples can be found at http://tinyurl.com/6gbjhd8 and http://tinyurl.com/67th5ga. (Present P&T policy permits only a two-page statement on teaching.)

Peer review of teaching often consists of two components, in-class observation of teaching, and a review of course materials, including student products. An online search of faculty handbooks provides the following information about practices at some of our peer institutions.

- Georgetown: peer review of teaching in P&T.
- Cornell: peer review of teaching materials in P&T.
- Northeastern: Peer review of teaching must take place at least once each year before promotion. A variety of inputs can be included, such as peer review of classroom performance and teaching materials; sample materials should be included in the dossier.
• Duke: uses just course evaluations from students.
• Stanford: Peer review, including classroom observation can optionally be included.
• UPenn: Letters from current students, former students, and teaching assistants must be included. Peer review of teaching can be optionally included.

We note that effective classroom review requires considerable training and time commitment by the peer faculty observers. Some committee members were concerned that the presence of a peer reviewer for summative assessment may distort the classroom sufficiently to lead to an unrepresentative situation. For these reasons, a majority of the committee did not believe we should recommend classroom observation for the summative peer review at this time. (Departments or Schools are free, of course, to require classroom peer review if they wish.)

**Conclusion:** Summative reviews for faculty should include peer review of the content and materials used in at least one course per academic year prior to the review. This review should be made by members of the department in which the teaching is taking place. The factors to be evaluated are indicated in the example rubric provided in Appendix A. Schools and departments should have the freedom to modify this rubric to suit their particular requirements. A number of these items could be satisfied by instituting a campus-wide required course description, web based, that includes the course syllabus, learning outcomes, assessments, etc. Such uniform descriptions would also be much more useful to students choosing courses than the current extremely brief descriptions in the General Announcements or the online course catalogue.

**Formative Peer Review**

While we refrain from recommending classroom peer review of teaching for summative purposes at this time, we strongly recommend that departments institute it in some form for formative review and faculty development. The goal is to have small faculty teams engage in discussions leading to the improvement of teaching, often by both the observed and the observer. The following procedure is suggested, suitably modified for particular departments.

Two faculty members agree to observe each other’s classes, for at least one course each academic year. Each makes two observations of the other, with a meeting after the first to discuss teaching and to design more specific goals/rubrics for the second set of observations. As a starting point, during the first observation, the reviewer should make free-form notes or use a simple check list to note if the instructor:
Was well prepared;
Seemed knowledgeable about subject matter;
Spoke clearly, audibly and confidently;
Used a variety of relevant illustrations/examples;
Made effective use of the board, visual aids and/or technology;
Asked stimulating and challenging questions;
Effectively held the class’s attention with active learning, and
Treated students with respect.
In addition, the observer should summarize what worked well in the class, and what could have been improved.

The department should keep a record that the reviews took place, but not the results or notes. The fact that the reviews took place should be part of the promotion and tenure dossier, but the outcomes of the reviews should only be used by the faculty members for self improvement. As faculty, departments, and schools become more familiar and comfortable with classroom peer reviews, and as people become trained in doing them, Rice may be able to move toward summative peer review of classroom teaching.

**Administration Response Needed**

Several WG members felt that efforts to improve the assessment of teaching effectiveness and course quality were hard to justify unless there were indications that the information would be used in a meaningful way. “If we build it, will they use it?” We feel that the administration needs to make strong and clear statements on several issues.

- Commit to changing promotion and salary decisions to include teaching effectiveness in a significant way, with much more weight than presently. Promotion materials should include student survey data, peer review data, and a longer teaching portfolio.
- Commit to eliminating the use of student comments in summative reviews.
- Provide resources and support to implement a more effective student course-instructor survey.
- Implement a policy requiring the peer review of course materials for summative reviews, and provide encouragement for formative classroom peer review.
• Require a common web-based course description that includes a meaningful syllabus, learning objectives, and the intellectual skills it hopes to provide (the eleven goals listed in the CRUP Report offer a good way to classify these).

• Undertake a program that recognizes and rewards service, making it also a meaningful part of the promotion process. As the university grows there is a greater need for people willing to spend time in faculty governance, committee work, evaluating teaching, mentoring, and a host of other tasks. This work is overburdening the modest number of people who see it as part of their basic academic duty even in the absence of recognition. There is growing conviction that only research and external funding, easily measured, receive significant recognition; a change of culture is needed.
## Appendix A: Example Rubric for Evaluation of Class Materials

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Exceeds expectations in all respects</th>
<th>Meets expectations in all respects</th>
<th>Meets expectations in most respects</th>
<th>Meets expectations in some respects</th>
<th>Meets expectations in few respects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course content includes appropriate topics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course content reflects the current state of the field</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course learning objectives are clear and appropriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course policies and rules are clear and appropriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture notes are well organized and clearly written</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplementary handouts and web pages are well organized and clearly written</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assignments are consistent with objectives and appropriately challenging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tests and exams are consistent with objectives and appropriately challenging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tests and exams are clearly written and reasonable in length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student products demonstrate satisfactory achievement of learning objectives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>