

Report of the Course Assessment Advisory Committee



**UNIVERSITY
AT ALBANY**
State University of New York

Fall 2012

Table of Contents

Executive Summary	1
Introduction.....	3
Charge 1: Summary of Historical Use and Policy	4
Charge 2: Formative and Summative Roles of Course Evaluations	6
Charge 3: Validity and Reliability of the Student Instructional Rating Form (SIRF)	9
Charge 4: Review and Recommend Changes to the SIRF	14
Charge 5: Usefulness of SIRF Summary Reports.....	17
Charge 6: Proposals for Increasing Student Participation in SIRF.....	18
Charge 7: Optional Items to Enhance Departmental and Faculty Use of SIRF	19
Charge 8: Student Access to SIRF	22
Charge 9: Gather and Review Reactions of Students and Faculty.....	25
Summary of Recommendations	26
References.....	30
Members of the Course Assessment Advisory Committee	32
Appendices A - G.....	33

Executive Summary

The University at Albany has a long history of collecting and using student feedback about courses and instructors, with such efforts dating back to the 1970s. Since that time, the University has periodically examined the design, administration, and appropriate uses of course and instructor evaluations produced by students. Recent national conversations about the importance of teaching and student learning, and what colleges and universities can do to improve both, combined with the University's move to an online course evaluation system in fall 2009, led Provost Phillips to establish the Course Assessment Advisory Committee in March 2010. The Committee was given nine specific charges around examining the history, design, administration, and appropriate uses of course and instructor evaluations produced by students.

Overall, based on its review of historical practice at UAlbany, the research and scholarship on course evaluations, current practice at other colleges and universities, and the importance of nurturing an engaging instructional environment, **the Committee believes that course evaluations, by themselves, provide useful but limited data for evaluating teaching, courses, or instructors.**

The accompanying report details the evidence and thinking the Committee considered in forming its specific recommendations around its charges. We were impressed by the thoughtful dialogue that informed our deliberations, as well as by the interest and comments of colleagues outside the Committee regarding this topic. The Committee's recommendations seek to:

- 1) Increase the usefulness of course evaluations to faculty and administrators by giving them improved tools to critique and improve upon teaching practices,
- 2) Acknowledge what course evaluations seem to gauge best – overall student impressions (a.k.a. satisfaction) with courses and instructors,
- 3) More fully engage students and increase their participation rates by proposing that selected course assessment results be systematically shared with students, under certain conditions, and
- 4) Highlight the need for other tools and techniques that assess and evaluate teaching better.

Major recommendations of the Committee in support of these objectives include:

- 1) The University should establish a mechanism to reinforce periodically its long-standing guidance in support of a multi-method approach to the assessment of teaching effectiveness. This recommendation lies at the heart of nurturing a long-term commitment to fostering instructional excellence.
- 2) University-wide summative course evaluations should include only a small number—currently two—of very general, global items, with the understanding that faculty members can choose to include additional items of their choosing.
- 3) A new teaching evaluation program, instrument, and reports, would be more useful to the institution, to faculty and their departments, and to students. Such a system would include:

- a. 2 global questions for students to rate their overall impression of the instructor and course,
 - b. 2-3 departmental questions: multiple choice survey items or open-ended text questions that the department sponsoring the course wishes to include for the purposes of formative assessment,
 - c. 2-5 instructor-designated multiple choice or open-text items survey items, the results of which would be restricted to viewing by the instructor.
 - d. A revised summary report.
- 4) UAlbany should join the growing number of institutions nation-wide in sharing selected evaluation results with students, but only for courses that achieve a 60% percent response rate.
 - 5) The Committee recommends that online courses, and perhaps discussion, laboratory, and other instructional settings that necessitate special or learning environments different from traditional lecture and seminars be assessed through departmental or instructor designed survey items.
 - 6) A standing Course Assessment Advisory Committee should be established, under the auspices of the Provost, with rotating membership from among the deans, department chairs, faculty governance, student leaders, and others involved in personnel decisions to provide ongoing guidance and feedback on the University's course and instructor assessment program.

In closing, we wish to thank the many individuals, governance councils, and various faculty and administrative bodies who provided constructive feedback and suggestions regarding our draft report. Broad-based campus participation and interest in the work of the Advisory Committee by the faculty is a testament to its dedication to UAlbany's academic mission; we trust that this dedication is reflected in the final report and recommendations.

Additional information about the work of the Course Assessment Advisory Committee, including historical policy documents that have guided UAlbany's approach to course evaluations since the 1970s is available on the Committee's wiki at <https://wiki.albany.edu/display/irpe/Course+Assessment+Advisory+Committee>.

Introduction

In March of 2010, the Provost convened the Course Assessment Advisory Committee to examine and make recommendations regarding a number of issues related to course evaluations. The Committee included department chairs, designees from the Council on Promotions and Continuing Appointments (CPCA), the University Planning and Policy Council (UPPC), faculty and staff with experience on the Council on Academic Assessment (CAA), the deans, undergraduate and graduate students, the Institute for Teaching, Learning and Academic Leadership (ITLAL), Institutional Research, Planning, and Effectiveness (IRPE), and representatives from the faculty at-large. Our charges were as follows:

- Charge 1. Overall, review the historical use of course evaluations at UAlbany and official policy regarding them, and to make recommendations regarding them going forward.
 - a. This includes use for professional development, as well as for tenure and promotion of the tenured and tenure-track faculty, as well as for how course evaluations are used to evaluate or promote the professional development of adjunct faculty.
- Charge 2. Specifically consider the formative and summative roles that course evaluations play (and what else beyond course evaluations should be used for both types of assessment), and make recommendations regarding this.
- Charge 3. Assess and make recommendations regarding the reliability and validity of the online course evaluation instrument.
- Charge 4. Review the questions/instrument currently in use (and for the past 20 years) and recommend changes/alterations, as the committee thinks fit to do.
- Charge 5. Review the usefulness of the summary reports for both formative and summative evaluation purposes, and make recommendations regarding them.
- Charge 6. Review various plans and proposals to promote higher student response rates in the online SIRF program, and recommend suggested action plans.
- Charge 7. Develop a bank of optional questions that faculty might choose from to add to their own course evaluations.
- Charge 8. Consider the desire of students to have access to course evaluation results, and the appropriateness of doing just that.
- Charge 9. Gather and review reactions of students and faculty to this past year's online evaluation experience to see where additional improvements can be made.

A list of Advisory Committee members, the Committee's formal charge, meeting minutes, and background and research materials are available for review at <https://wiki.albany.edu/display/irpe/Course+Assessment+Advisory+Committee>

The following terms are used throughout this document:

- “Course evaluation” is the instrument used by students to provide their impressions of the quality of a course and an instructor.
- “Evaluation of teaching” is the comprehensive process of considering data of various types and sources in the assessment of an instructor’s overall teaching skills.
- An “item” is an individual question or prompt on a test or survey (e.g., “Was well-prepared for class”), as well as the choice options (e.g., “Almost never/seldom/as often as not/very often/almost always”).
- “Summative evaluation” is the use of any test, survey, or other systematic gathering of information at the *end* of a specified time period of interest (e.g., unit, semester, school year), and which is primarily used to categorize the performance of an instructor. Its main purpose is to obtain a measurement of performance which is to be used in decision making or classifying instructors according to defined performance categories (e.g., *Basic*, *Proficient*, or *Advanced*). Because this primary purpose requires high quality information about *overall* performance, summative evaluations are typically designed to yield highly reliable and valid total scores. Summative evaluations are usually course-grained. As such, they are typically not intended—and are thus less well-suited—to provide individual diagnostic information about instructors, nor to yield recommendations for improvement in learning or teaching (Cizek, 2010).
- “Formative assessment” is a system of information gathering the purpose of which is to identify an instructor’s strengths and weaknesses and generate recommendations for improvement (Cizek, 2010). Formative assessment information tends to be confidential: Only the instructor has access to it. Although formative assessments are generally administered midstream, end-of-semester assessments may also be used formatively.

Charge 1: Summary of Historical Use and Policy

There is a long history at the University at Albany of student and peer assessment of teaching. Assessment of teaching quality and effectiveness is a major factor in reappointment, tenure, and promotion at the University at Albany but the scope and nature of this assessment has changed over time. Currently, Academic Affairs’ guidelines for consideration of faculty for continuing appointment and promotion mandate that both peer and student evaluations must be presented as part of the case.

The most important legislation governing assessment of teaching is Senate Bill 8384-07, implemented by administrative memorandum in April of 1984 and revised in 1991. By policy, student feedback is regularly solicited for courses, using the Student Instructional Rating Form (SIRF) coordinated by the Office of Institutional Research, or some other instrument endorsed by the instructor’s department or program. The guidelines promulgated in 1984 specify that “all students shall be given an opportunity to make an evaluation in every class each term” and mandate that the collection of student opinions should be formulated and administered systematically at the department level (2005-06, p. 2). This system was developed to improve the coherence and utility of information on teaching in support of personnel recommendations, first for full-time faculty and later for part-time faculty. The SIRF system was developed at a time

(early 1980s) when the Student Association assessed teaching through its own system, to assist students in course selection.

The 1984 Senate legislation went beyond student course evaluations. It also called for peer evaluation of teaching and noted that the methods for proper peer review had been less fully considered. To support peer evaluation of teaching, the legislation described several accepted techniques as examples for the faculty of each unit to use in developing a system that is tailored to the particular needs of their curriculum. It also mandated the use of peer evaluation in decisions concerning continuing appointment, but noted that departments were to be given “broad latitude” in developing systems for collecting and interpreting peer evaluations of teaching (2005-06, p. 1). Such peer evaluation had both formative and summative purposes, such as improving teaching and providing information for personnel actions. The operative policy now does not demand any particular form for peer evaluation and explicitly leaves open the question of whether peer classroom observation should be incorporated, but it does demand that departments “...establish a credible and defensible method of evaluation of teaching”; evaluations must implicitly recognize the department’s expectations concerning teaching and “...should provide both a judgment of the instructor’s competence and an explanation of how that judgment was arrived at” (2005-06, p. 7).

The current version of the SIRF asks students to rate their instructors and courses on a 5-point scale (i.e., almost never, seldom, as often as not, very often, almost always; and includes an option for don’t know/doesn’t apply). Seven SIRF items, listed below, regard how frequently the instructor:

1. Was well-prepared for class.
2. Communicated course content in ways you understood.
3. Stimulated your interest in the course material.
4. Challenged you intellectually.
5. Was receptive to students' ideas and viewpoints.
6. Was available outside class to discuss course matters.
7. Held you to high standards of performance.

Two additional items ask students for overall ratings:

8. Instructor, overall: Poor, fair, average, good, excellent.
9. Course, overall: Poor, fair, average, good, excellent.

In 2008-09, after a one-year pilot program, ITLAL introduced a confidential, mid-term student survey designed to provide formative feedback on teaching. This has subsequently been used by an increasing number of faculty. In the spring of 2009, the University conducted a pilot program for administering the SIRFs online, and in fall 2009 switched the SIRF program fully to an electronic form. This switch also allows the use of individualized departmental assessments and does not preclude departments from using their own instruments. However, the response rates of this newest form of implementing the SIRFs have been disappointingly low compared to in-class administration (an average response rate of 74% in-class paper administration at the Lower

Division level versus 47% online; 78% in-class at the Upper Division level versus 50% online; and 87% in-class versus 68% at the Graduate level¹).

Currently the Student Association does not gather or disseminate information on teaching, but Student Association leaders in 2009 initiated conversations with the Provost's Office to explore doing so. Proprietary online services, particularly ratemyprofessors.com, do collect ratings of University at Albany faculty and make them available to students and the public. The Student Association has explored ratemyprofessors.com and similar products as means to help inform student course selection. For a number of reasons, including limited sampling and insufficient or even inappropriate survey items, the Committee believes that such services are not ideal for helping students select courses.

Student course evaluations, and the SIRF in particular, are also used to evaluate and guide the professional development of adjunct faculty. University [Guidelines for the Appointment of Part-Time Faculty Members](#) stipulate that:

“The University requires all academic units to establish procedures for the evaluation of teaching faculty, including part-time faculty. Insofar as possible, the procedures should parallel those used for full-time faculty members, and at a minimum should include provision for student course evaluations and review of course syllabi and related materials.”

Part-time faculty are also encouraged to take advantage of the professional development opportunities, workshops, seminars, and other programs offered through the Institute for Teaching, Learning and Academic Leadership (ITLAL), Academic Computing, the Office of Academic Affairs, and other units. Part-time faculty course evaluation results may be used by individual faculty to assist their pursuit of these developmental activities.

Charge 2: Formative and Summative Roles of Course Evaluations

We believe that course evaluations provide useful but limited data for both summative and formative evaluation of teaching. As a single measure or even as one of two measures, course evaluations are insufficient evidence to uphold the integrity of the evaluation process. Teaching is a complex enterprise, and effective instructors continually test new methods, techniques, formats and assignments. The summative evaluation of teaching must account for the value of innovation and for the complex, explorative, risk-taking process that innovation requires. If course evaluations are the most heavily weighted measure of teaching, instructors will have low incentive to take sustained risks to adopt new approaches that might lower course evaluation scores during the experimental/developmental phase of an innovation.

¹ These figures derived from the SIRF system for all participating sections, in the aggregate, from fall 2005 through spring 2010.

Summative Use of Course Evaluations

For the summative evaluation of teaching, research shows that, in general, scores for global items (e.g., “Overall effectiveness of course” and “Overall effectiveness of instructor”) on course evaluations correlate highly with instructor peer evaluations of the same course (Cohen, 1981; Feldman, 1998; Murphy, 1987). Broad questions that ask students to score the general value of a course work for this purpose. Reinforced by the analyses below, **we recommend that global items used in university-wide summative course evaluations continue to be held to a small number of very general items—currently two—with the understanding that a faculty member can choose to include additional items in tenure and promotion dossiers.** However, because evaluation forms that ask little more than whether the student liked a course and thought highly of the instructor run the risk of measuring personality and charisma more than the ability to promote learning, we also recommend that more specific items related to learning, specific instructor behaviors, and to the student experience be used for summative purposes when they reflect the values and interests of particular instructors, departments and/or colleges.

More specific analytical questions can be problematic when used for summative purposes. Analytical questions capture details that may be formatively useful to the instructor but may not be indicative of the quality of the overall student experience, or are inappropriate for evaluation of teaching in the context of various teaching purposes, formats or methods. For example, “Was available outside class” is not a transparently valid item in the case of online or blended instruction, nor in the case of collaborative classrooms where student-instructor interactions in the classroom are constant. “Challenged you intellectually” or “Was receptive to students’ ideas and viewpoints,” while highly desirable qualities overall, might not be fully consistent with the goals of a required basic technical skills course. Developing a particular skill may or may not carry with it a high intellectual challenge or require conditions for multiple viewpoints. **We therefore recommend that University-wide results summaries used for evaluation purposes do not include these more specific analytical questions, as their applicability across instructional settings is highly variable.** Analytical items of the instructor’s and/or department’s choosing should remain in instructor and/or department summaries for formative assessment purposes.

Formative Use of Course Evaluations

For formative evaluation of teaching, course evaluations play an essential and legitimate role of providing instructor feedback from the students’ perspective. This role is best supported under three conditions:

1. When the feedback is collected and returned to the instructor in a timely manner.

Information from course evaluations that arrives well after the instructor’s next teaching assignment begins has limited use—whatever its content—since the data will not help inform habits of planning and design.

2. When the course evaluation instrument aligns with the specific goals and practices of the instructor. Because teaching is highly individualized by method, level, format, discipline and other factors, non-traditional, innovative, or alternative teaching practices

may be difficult for students to assess adequately using a standard evaluation instrument that does not account for these many critical variables.

3. When an instructor has the ability to insert individualized analytical items into the course evaluation instrument, and can receive the results confidentially (i.e., results are not included in summative evaluation). An instructor who chooses to measure a particular element of teaching will need the freedom to ask questions that may provide useful feedback but may not necessarily reflect positively on perceptions of his/her teaching performance.

The University has made strides towards enabling instructors to use course evaluations more formatively in recent years. Beginning in spring 2006 departmental SIRF averages, at the lower, upper, and graduate division, are included with each instructor's class summary reports to foster self-reflection about the past semester's teaching experience. Previously, departments received the summary reports. Some passed along these reference data to their instructors, some did not. Sharing summary reports directly with instructors was done at the urging of the Council on Academic Assessment. Additionally, the online SIRF system now automatically delivers evaluation result summaries to the e-mail in-boxes of instructors as soon as the semester's grades are posted, up to two months before the prior in-class paper SIRFs were often processed. Instructors now can also retrieve their online result summaries on-demand via MyUAlbany.

Recommendations Regarding Formative and Summative Uses of Course Evaluations

For Summative Evaluation:

- Develop and publish examples of possible evidence for the evaluation of teaching. The examples would help faculty members identify the types of evidence that could be included in a teaching dossier;
- Consider a transitional framework that allows faculty currently in the tenure pipeline to build upon their existing strategies for assessing teaching as the university moves towards a more comprehensive model of assessment. This might be accomplished by utilizing a narrative structure, or teaching portfolio (see Selden, 1991; Urbach, 1992), that allows faculty to describe the scope and quality of their teaching through structured reflection and thoughtfully selected information about their teaching activities, providing evidence of instructional effectiveness;
- Develop one or more tools or processes that can be adapted across disciplines to guide administrators' evaluation of a faculty member's teaching dossier;
- Emphasize that peer assessment of teaching may include classroom visitation and observation, but is most useful when encompassing a broad range of activities related to the assessment of teaching by one's colleagues. These include, but are not necessarily limited to, the review of course planning and design (e.g., syllabus, Web presence), instructional materials (e.g., handouts, exercises, readings, lectures, activities), and assessments (e.g., tests, graded assignments), as well as in-class interaction with students and instructor presentations.

- Restructure the SIRF to limit the number of global items to two, while expanding the flexibility of administrators and individual faculty to customize the form for specific evaluation goals;
- Expand the Likert-type scale used on the survey instrument to seven categories from the five currently in use, as the Committee believes this would capture more variability in the student response set, and would therefore be more useful to faculty.²
- Make more explicit during the hiring process the expectations for effective teaching, and share with new hires the written standards and instruments for evaluation of teaching.

For Formative Assessment

- Build (through policy) an expectation and a support process for regular formative assessment of teaching for faculty, with a shorter frequency (1- 2 years) for pre-tenure faculty. The formative assessment would include a brief peer review of evidence, including evidence from different sources, determined on a case-by-case basis to include classroom visit or video clip review; analysis of course evaluations; review of course materials; etc.
- Develop and publish (via ITLAL) examples of how faculty in various disciplines can structure narratives of teaching success, and incorporate them in the tenure and promotion process for tenure-track faculty, and in the review and reappointment processes for contingent faculty.

Charge 3: Validity and Reliability of the Student Instructional Rating Form (SIRF)

The Office of Institutional Research, Planning, and Effectiveness (IRPE) constructed a data set of all Student Instructional Rating Form responses for fall and spring semesters from fall 2005 through spring 2010. In all, the data set consisted of 319,320 cases, one record for each student evaluation that was processed for each class each semester. Student ratings were captured anonymously in the data set. Individual instructors (and their rank) may be tracked across semesters in the data set using a randomly assigned identifier; and no identifying information about instructors is contained in the data set. The number of cases available for statistical testing varies from 283,646 to 319,320 depending on the specific measure of interest and the degree of missing data to which it is subject.

Preliminary Examination of Item Means. Statistical testing suggested that the first seven instructor-related items largely reflected students' perceptions of the overall instructor and course items. In other words, students did not appreciably differentiate between their overall ratings (items 8 and 9) and their ratings of more specific characteristics of the instructor and course (items 1 – 7). For this reason, our primary analyses concentrated on the two items assessing overall instructor performance (item 8; “Instructor, overall”) and overall course

² While Pajares, Harteley, and Valiante (2001) used a 100 point Likert-type scale to increase prediction, the consensus view of the Committee was that seven response categories would be sufficient.

evaluation (item 9; “Course, overall”). The overall average rating for instructor was 4.20 (N = 317,417); the average score for course was 3.94 (N = 316,023).

Traditional Evaluation (in-class) versus Online Evaluation. The in-class ratings were slightly higher than the online evaluations for both instructor (4.22 versus 4.10) and course (3.95 versus 3.89).³ Because the patterns for all other variables described below were comparable between both types of evaluation, online and in-class scores were combined. See Appendix A: Base Statistics and Rating Breakdowns by Various Student, Course, or Instructor Attributes.

Instructor Rank. Although distinguished professors received the highest scores for both the instructor and course items, there was no clear pattern between different instructor ranks.

Student and Instructor Gender. Female students rated their female instructors slightly higher and their male instructors slightly lower for the instructor and course items. Male students rated their female instructors slightly lower and their male instructors slightly higher for the instructor and course items. However, the boost to the “Instructor, overall” rating from having students of the instructor’s own gender respond to the survey was very small (0.03), regardless of whether the instructor was male or female. At the aggregate level both male and female instructors matched the “Instructor, overall” item average of 4.20, on a scale of 1.0 to 5.0.

Major/Minor. Students taking a course for their major or minor rated their instructor and the course almost identically. Students in courses outside of their majors or minors rated both the instructor and course slightly lower than did students in their majors or minors.

Time of Day. There was an interesting relationship between time of day and student ratings. Ratings were lowest for both instructor (4.14) and course (3.88) for classes scheduled during midday. Ratings were highest for the earliest and latest classes.

Size of the Class. Class size was related to instructor ratings, with the smallest classes (less than 20) rating their instructor the highest and the largest classes rating the instructor the lowest. The ratings for the course item, although lower, followed the same pattern, with the smallest classes assigning the most positive ratings and all subsequent class sizes assigning lower ratings as class size increased.

Course Level. There is a general rise in both instructor and course ratings as course level rose from lower level to graduate level. However, when holding constant student level (e.g., freshmen, sophomore, etc.) or class size, virtually all of the difference in ratings vanishes. Thus, the effect for course level appears to be a function of student level and class size.

Self-Reported GPA and Self-Reported Grade. Self-reported GPA was directly related to both instructor and course evaluations; students with higher GPAs gave higher ratings. There was a similar relationship regarding self-reported expected grades, except that students expected to

³ Statistical significance levels are not reported because the data set used in the analysis covers the entire population of student ratings, rather than a sample, and therefore the probability that the observed means are unlikely to have occurred by chance is 100 percent.

receive an “E” or “U” rated both the instructor and the course higher than students expecting to get a “D.” GPA categorized by class was also analyzed. In general, there was fairly weak but nonetheless present relationship between class GPA and these ratings; as class GPA rose, so did the ratings.

Instruction Type. Many different instruction types (e.g., lecture, discussion) were investigated. Students in lecture classes provided the lowest ratings for the instructor and course evaluation items. Thesis and contract courses were the highest rated.

Regression Analysis

We investigated the extent to which the above factors influenced the “Instructor, overall” item using regression analysis. Regression allows us to determine to what extent several factors together influence another variable, such as instructor rating. Given that regression works best with relatively continuous variables, such as the number of students in a class, versus a variable with many categories, such as instruction type, we did not include all of the factors discussed above in the regression equation. We included student level, expected grade, total enrollment, whether the class was a requirement for the major or minor, average GPA for the class, evaluation response rate for the class, time of day the class met, level of the class (i.e., lower division, upper division, or graduate), instructor rank, student gender, and instructor gender in two regression models, one for traditional evaluations and one for online evaluations. Nominal level variables were dummy coded.

These independent measures explained about 10 percent of the variance in the overall instructor rating. Together, they accounted for 10% of the variance (model R-Square) in the instructor rating for traditional evaluations and 13% of the variance (model R-Square) for online evaluations. The single biggest contributing measure for both online and traditional evaluations was the student’s expected grade, as evidenced by standardized regression coefficients. The next biggest contributor was average GPA for all students in the class. The square of the partial correlation coefficient indicates that expected grade explains 4% of the total observed variance in the model for traditional in-class evaluations, and 7% of the total observed variance in the model for online evaluation results, controlling for all other measures.

However, it should be noted that all of these analyses are based on data at the student level. It is also informative, perhaps even more informative, to investigate the average rating that each instructor receives from students in his/her class (i.e., analyze class-level data) rather than only investigating students’ ratings of their instructors (i.e., analyze student-level data). Using data from one semester only (Spring 2010) because of the size of the data set, we first determined whether students in a class display reasonable levels of agreement in how they rate their instructor to warrant aggregation of the student ratings to the class level. The two most commonly used indicators of agreement, intraclass correlation coefficients (ICC)(1) and ICC(2) were very good [ICC(1)=.27 and ICC(2)=.84] indicating substantial agreement in instructor ratings. Therefore, class-level scores on relevant variables (e.g., average instructor rating) were computed and regressed onto class characteristics.

The results suggest that the class characteristics discussed above have even stronger effects when viewed on the class level, which arguably provides a more accurate estimation of the influence of

these variables. The most important variable is the students' expected grade. The individual level data suggests that the correlation of expected grade with instructor rating is .26 (explaining about 7% of the variance in instructor ratings). However, when the data is aggregated to the class level, the correlation between expected grade and instructor rating is .43, accounting for approximately 18% of the variance in instructor ratings.

A regression was also run using the variables expected grade, total enrollment, average GPA for the class, evaluation response rate for the class, and whether the instructor was a TA. The combination of variables accounted for 20% of the variance, in comparison to the 10% of variance accounted for by the regression on the level of the individual student data that included these variables along with others described above. The class GPA and class enrollment also have small but substantial effects. Thus, it appears that class characteristics have even stronger effects when investigated at the class-level than the student-level.

Interestingly, the [1982 SIRF Pilot Study](#) (p. 31) found the correlation between students' expected grade and overall instructor rating explained only eight tenths of one percent of observed variance in comparison to the 18% observed in the class level data. The relationship between students' expected (or actual) grade and their ratings of instructors are potentially of interest in terms of the validity of the ratings. Some may argue that students reward faculty for lenient grading (or retaliate with overly negative evaluations if their work is rigorously graded), while others might argue that good teaching leads to good learning, which leads to better grades and better ratings. The committee believes that the answer probably lies someplace in the middle, and *recommends that the University reinforces its long-standing support of a multi-method approach to the assessment of teaching effectiveness.*

Appendix B contains the bivariate correlations and full regression model results for the in-class and online evaluations, with the instructor overall rating as the criterion measure both on the individual level and the class level.

Factor Structure of the Student Instructional Rating Form (SIRF)

We examined item correlations, analyzed factor structure using exploratory factor analysis, and measured reliability using Cronbach's alpha. **All analyses point to the same conclusion: students are rating all items mostly according to a global reaction to the instructor and are not differentiating much, if at all, among the specific items, regardless of content.**

Item Correlations. We examined the correlations among the two overall items designed to assess students' overall impressions of the course and instructor (items 8 and 9) and the seven items on SIRF designed to assess specific instructor competencies:

1. Was well-prepared for class
2. Communicated course content in ways you understood
3. Stimulated your interest in the course material
4. Challenged you intellectually
5. Was receptive to students' ideas and viewpoints
6. Was available outside class to discuss course matters
7. Held you to high standards of performance

We examined the correlations separately for traditional (in-class) and online evaluations. The seven items designed to assess instructor competencies all correlated very highly with each other. Appendix C contains the bivariate correlation results for each administration mode; in-class and online. Even the lowest correlations (.49 for both the correlations between item 6 and item 1, and item 6 and item 4 for traditional evaluations, and .52 for the correlation between item 1 and item 3 for online evaluations) represent moderately large correlations. The largest correlation was between item 2 and item 3 (.74 for traditional and .78 for online).

The seven instructor competency items also correlated strongly with both the overall course rating (item 9) and the overall instructor rating (item 8). Examining the correlations with instructor for the traditional evaluations, the strongest relationship was with item 3 (.78) and the only correlations below .6 were with item 1 (.597), item 6 (.567), and item 7 (.595). Overall, the correlations between the seven items and the overall course rating were similar to those of the overall instructor rating. We also examined the online evaluations and found very similar relationships as found in the traditional evaluations: The same items correlated lowest and highest. Lastly, it should be noted that the overall course (item 9) and overall instructor (item 8) ratings were strongly correlated for both traditional (.78) and online (.78) evaluations. Correlations of .85 or greater are generally viewed as an indication that raters may be not distinguishing between items.

In summary, the patterns of results were very similar for online and traditional evaluations. For both types of evaluations, all item correlations are .48 or above. Furthermore, high correlations between overall instructor and course ratings suggest that students are not distinguishing between their overall evaluation for the course and for the instructor.

Exploratory factor analysis. We conducted exploratory factor analysis using maximum likelihood and oblimin rotation, in case more than one factor was indicated. Exploratory factor analysis is a statistical technique which allows us to gain a better sense of the number of constructs (e.g., words or ideas that underlie scores such as ratings). For example, it may be the case that, when filling out the nine items on the evaluation sheet, students' answers to four of the items reflect their general impression of the instructor and their responses to the other five items reflect their general impression of the course, regardless of the wording of the items. In this example, factor analysis would suggest that there are two factors underlying student responses to the evaluation items.

We conducted the factor analysis separately for both online and traditional evaluation methods and also for the seven instructor competency items, as well as with the seven instructor competency items plus the two overall items (course and instructor). In all cases, only one factor with an eigenvalue greater than 1 emerged. This factor accounts for the majority of the variance (seven items, 63% traditional and 67% online; nine items, 65% traditional and 67% online). We also examined the factor loadings, which were all substantially above the .30 cut-off used to indicate a "good" factor loading. In fact, the lowest loading was for item 1 (.67 for traditional and .69 for online). The highest loading item was the overall instructor item (.90 for traditional and .89 for online). Appendix D contains the factor analyses for all four models examined.

We also examined whether the factor structure differed according to expected grade (collapsing traditional and online) and found that, for those students expecting to fail a class, a one factor solution accounted for 71% of the variance, which is much more variance than the one factor

solution accounted for in regard to the other expected grade categories (57% for “C”, 59% for “B”, and 59% for A). Thus, it appears that failing students are even more likely to use a global evaluation of the course when answering all items. In general, the above analyses suggest that students are basing their item ratings mainly on a general impression of the instructor and are not differentiating among the seven items. These findings suggest that the items do not serve formative purposes by providing feedback to instructors regarding their instructional techniques.

Reliability

Cronbach’s alpha was used to assess the extent to which the items correlate with each other. Measures with an alpha of .80 or greater are typically considered reliable. Cronbach’s alpha was separately calculated for SIRF data gathered traditionally in-class and online, across the seven instructor competency items (.90 traditional and .92 online) and the seven competency items plus the two overall items (.93 traditional and .94 online). The analyses reveal a high degree of internal consistency among the items. Appendix E displays the results of the reliability tests. These results further reinforce the observations above that the seven instructor sub-items are measuring the more global constructs of instructor and course overall ratings.

Response rates

The validity, reliability, and representativeness of any survey can be threatened by low response rates. The traditional SIRFs had a footnote that read: “Caution is advised if the number of non-responses is large with respect to the number of respondents.” Given the uniformly lower response rates for online SIRFs, ***the committee proposes that online SIRF ratings for any class that has a response rate below 30% be viewed with caution.***

Charge 4: Review and Recommend Changes to the SIRF

In addition to the statistical observations detailed above, the Committee compared the UAlbany SIRF items to items on course evaluations from seventeen different colleges and universities⁴. A convenience sample, derived from internet searches for university evaluation forms and from those universities on UAlbany’s peer institution listing, was the basis for the comparison group. The comparisons suggest that the UAlbany items are broadly consistent with those used on other evaluation forms, although the nine SIRF items do not all have cognate members on each of the seventeen course evaluations examined. Table 1 below lists the nine UAlbany SIRF items, along with the percentage of the seventeen course evaluations that contained a corresponding version to that item.

Table 1.

⁴ Course/instructor evaluation forms were found online at: Brandeis University, Cornell University, Lehigh University, North Carolina State University, Northern Illinois University, Northwestern University, Princeton University, Syracuse University, University of Arizona, University of California at San Diego, University of California at Santa Cruz, University of Colorado, University of Iowa, University of Minnesota, University of Oregon, University of Virginia, and the University of Wisconsin at Milwaukee - School of Business Administration

SIRF Content Representation in Seventeen Other Institution's Course Evaluations

SIRF Item	% of Sample with Similar Item
1. Well prepared for class	59%
2. Communicated course content in ways you understood	59%
3. Stimulated interest in the course material	41%
4. Challenged you intellectually	41%
5. Receptive to students' ideas and viewpoints	47%
6. Available outside class to discuss course matters	65%
7. Held you to high standards of performance	12%
8. Instructor Overall	76%
9. Course Overall	76%

With one exception, a version of each of the nine SIRF items appeared in these questionnaires at least 41% of the time. The exception is item 7. Versions of this item appeared in only two of the seventeen questionnaires examined. Item 7 may be less frequently represented in the sample of questionnaires due to the more frequent occurrence of item 4 in the versions of the course evaluations examined. It can be argued that there is some conceptual overlap between items 4 and 7. The strong representation of UAlbany SIRF-like items on other institutions' evaluation forms suggests that the UAlbany SIRF is in the mainstream of course evaluation instruments used by other schools.

While UAlbany's SIRF items are similar in theme to those in use at other colleges and universities, the Committee's regression, reliability, and factor analytic analyses above suggest that these various items are essentially measuring more global instructor and course dimensions. Although these items are widely used, we believe them to add little value, either formatively to instructors, or summatively to the administration.

If the purpose of the SIRF items is to provide instructional feedback to instructors, it is difficult to recommend any of the items in their current form, given that all appear to be strongly related to an overall instructor or course impression, as discussed above under Charge 3. While there are small differences in item correlations among the SIRF items, and some rank higher or lower than others, in an absolute sense, all correlations among the core SIRF items represented strong relationships. Thus, ***our recommendation is that, for the purpose of providing information that is not a function of students' overall impression, none of the items should be used in their current form.***

It may be possible to improve the items for formative use with some relatively minor format changes. The performance appraisal literature shows that, by providing behavioral descriptors to both the items and to the rating scale, raters may gain a better understanding of both the item and the response options, resulting in more high quality feedback. It is worth noting that, of the current items, the more behavioral items (items that reference specific teaching behaviors) appear to be least influenced by students' general impression of the instructor, as exemplified by bivariate correlations in the .60 and .57 range, as opposed to other items which exhibit bivariate

correlations above .60, and as high as .78. We also note that item 3 (“stimulated your interest”), one of the most abstract items, appears to be one of the items most influenced by students’ impression of the instructor, as its bivariate correlation with overall impression of the instructor, at .78, is highest among these items. Thus, it may be worthwhile to add some behavioral descriptors to both the items and the response options, making it easier for students to distinguish between items.

Lastly, even though there is a clear theoretical distinction between course and instructor evaluations, this distinction is not represented in the student ratings. ***We recommend that, if the purpose of the ratings is to gain a global sense of instructor performance as evaluated by students, that the overall instructor item (item 8) should be used, as it captures this global impression well.*** The exploratory factor analysis for both online and traditional evaluations showed that this item had the largest loading on the underlying factor. However, one item measures are generally not recommended; perhaps additional overall instructor items should be considered, or the overall course rating should be retained.

The Committee recommends that a new teaching evaluation instrument and program, which will employ up-to-date teaching assessment methodologies, be used for multiple functions:

- to provide consistent information on teaching in support of personnel recommendations for tenure track and tenured faculty;
- to improve teaching practice and professional development of full-time and part-time faculty; and
- to help students in course selection (see Charge 8, below).

We recommend the adoption of an assessment system similar to the one employed by the University of North Carolina, in which assessment items are matched to one of four user groups for purposes of functionality and levels of access: (1) Institutional research purposes, university-level promotion committees, and other system administrators, (2) deans or department chairs, (3) instructors, and (4) students.

A small number (2-5) of global measures should be used at the university level. Departments should add their own items (1-5), which would remain at the department level or be shared at the university level at the department’s discretion. The instructor should be able to select (1-5) items that are used exclusively for formative purposes and which are not released to other user groups without the instructor’s permission. Open-ended text questions may also be eligible for instructor or departmental items. Finally, items identified by student interest groups and approved by instructors should be released to students for the purpose of making course selection decisions. UNC’s functionality grid can be found in Appendix F.

An online menu of formative assessment items should be made available for instructors to select for each semester course. UNC’s item bank can be found in Appendix F. We recommend refining this list and developing meaningful conceptual categories that would allow instructors to locate and select items easily.

Based on the statistical analyses noted earlier which found that the seven instructor sub-items largely reflect a global dimension of the instructor and/or course, and what appear to be “best

practice” at institutions like the University of North Carolina, *the Committee recommends that the UAlbany SIRF include:*

- **2 global questions for students to rate their overall impression of the instructor and course.** These items would be used for summative evaluation of the instructor by departments, deans, and the Council on Promotions and Continuing Appointments (CPCA), and would also be shared with students as an aid to course selection.
- **2-3 departmental questions which are multiple choice survey items or open-ended text questions that the department sponsoring the course wishes included for purposes of formative assessment.** These items could be developed internally by the department, or selected from items already in use at other institutions (e.g., the UNC item bank). These items may or may not be part of faculty tenure and promotion portfolios. Departments, as they do now, could continue to include open-ended text questions of their own choosing on SIRFs for their courses.
- **2-5 instructor designated multiple choice or open-text items survey items.** Item summaries would be restricted to the sole province of the instructor. It is hoped that the instructor will communicate back to students in succeeding semesters how he/she uses this information to enhance or modify the course. It is suggested that instructors wishing to develop more broad-based diagnostic items make use of ITLAL’s confidential, mid-term student survey, which is designed to provide formative feedback on teaching directly to instructors.

In addition, *the Committee recommends that students be asked to select the instructor’s name from a dropdown menu in each evaluation as a reminder of which instructor they are rating at any particular time.*

Appendix G illustrates what a new UAlbany SIRF form might look like under this proposed design structure.

We recommend that a standing Course Assessment Advisory Committee be charged to develop guidelines for how instructor and departmental survey items should be formatted (e.g., scale construction, item stem and response categorization, etc.) and what the approval process will be, if any. For example, would items be chosen from an approved list, or should the University allow great latitude to instructors and departments? While our inclination is to allow broad latitude to instructors and departments, some type of centralized oversight/approval seems prudent.

Charge 5: Usefulness of SIRF Summary Reports

In addition to aggregate data on student characteristics such as GPA, expected course grade, gender, etc., the current SIRF summary report provides item-specific information about the frequency of ratings, means and standard deviations. Although department chairs and others on the committee report finding the item-specific information useful, the analyses reported above raise questions about the validity and usefulness of the ratings data. *We recommend that the*

summary reports be revised as the SIRF itself is revised, and that the feedback from the university community regarding this report also be incorporated into future report design.

Charge 6: Proposals for Increasing Student Participation in SIRF

UAlbany online SIRF student response rates are in the general range of those experienced by many other colleges and universities, and both the higher education literature (Avery, et al., 2006; Dommeyer et al., 2004) and the Committee's research suggest that overall ratings results vary little between online and in-class paper evaluations. Nevertheless, the very fact that the response rates are lower than instructors' experience with in-class evaluations leads to reasonable faculty concern that online results may be skewed negatively, or positively, unless a sizable majority of the class participates. To allay this concern, the Committee has considered sharing evaluation data regarding the course and instructor overall survey items with students in order to increase participation in the online evaluation process and thereby to generate potentially more meaningful data and feedback (for a full discussion of the wisdom and implications of sharing evaluation data with students, see below under Charge 8). Other institutions have taken this step, and their experiences and the Committee's deliberations have suggested a range of options to stimulate participation, from draconian to softer incentives, such as:

1. Share the course and overall SIRF results only for courses that achieve a 60% percentage response rate.
2. Individual students gain access to data only if they have filled out evaluations for all of their courses during the previous semester.
3. Reward students who complete evaluations with early registration, and/or early access to semester grades.
4. Enter students who complete evaluations in lotteries for bookstore gift certificates, dining hall credits, etc.
5. Restrict all access to grades in MyUAlbany for those students who do not complete all of their evaluations.

The Committee recommends the first option listed above. Providing access to the overall course and instructor evaluation results only if a sizable percentage (60%) of students in each class complete their surveys seems reasonable, and strikes a sensible bargain with students—take the evaluations seriously, fill them out, and the University will share key results with you, the student.

We also believe that providing students with information that they actually want will also increase their incentive to participate in SIRF. The student intersession survey, as discussed below under Charge 8, identified several informational items students preferred to have made available, including: “The instructor explained complex material clearly;” “The instructor did a good job of covering the course objectives and content;” “Course expectations and requirements overall;” and “The degree to which examinations and quizzes (and other evaluations) covered the

course content and objectives.” Our earlier analysis suggests that most SIRF survey items are conflated with the course and instructor overall items. ***We recommend that overall course and instructor evaluation results be shared with students in order to increase student participation and the usefulness of evaluations to students.***

We recommend further that instructors share with students how they utilize SIRF results, both the global items and the instructor and departmentally designed items. This information would be most valuable to share with students just prior to launching each semester’s SIRF evaluations, but is also worth sharing with students whenever curricular or pedagogical matters are discussed, since this can also serve to foster student engagement more generally. ***The Committee also recommends that each department reinforce this communication stream with its tenure-track and adjunct faculty within its teaching policies, and that a new standing Course Assessment Advisory Committee address the SIRF program administration protocol that instructors and departments follow.***

Charge 7: Optional Items to Enhance Departmental and Faculty Use of SIRF

Listed below are items from other course evaluations that might productively expand upon the set of items the UAlbany SIRF currently uses. In some cases these items add a new dimension. In others, they elaborate on themes that have surfaced in recent campus-wide discussions (e.g., General Education Taskforce, First-year Experience Taskforce). Some of the items selected have overlapping content because they are from different questionnaires. We list them in Table 2 to illustrate the diverse interests they might serve, and which are perhaps idiosyncratic to particular instructors or courses.

Table 2.

List of Potential Optional Items

Instructor-Related Items

- 1) The instructor invites criticism of his/her own ideas.
- 2) This instructor creates an atmosphere where ideas can be exchanged freely and easily.
- 3) I am encouraged to participate in the classroom.
- 4) Course material is presented enthusiastically.
- 5) The instructor seems interested in teaching this course.
- 6) Student questions are encouraged.
- 7) Help is available outside class if I have questions.
- 8) The graded assignments allowed me to demonstrate what I learned in the course.
- 9) I received feedback that helped me see ways in which I could improve my learning and

understanding.

- 10) The instructor never intimidated or embarrassed students.
- 11) The instructor returned my work in a reasonable time.
- 12) The instructor graded in a consistent and systematic way.
- 13) The instructor prepared a syllabus that accurately reflected what was expected of the student.
- 14) The instructor demonstrated a thorough knowledge of the subject matter.
- 15) Classroom instruction holds your attention.
- 16) Instructor's style facilitates understanding responses and note-taking.
- 17) Instructor promotes appropriate questions/discussions.
- 18) Graded work is representative of course material.
- 19) Rate how well the instructor encouraged class participation and discussion.
- 20) Rate how well and timely the instructor provided valuable feedback on graded material.
- 21) Rate the instructor's enthusiasm about the subject.

Course-Related Items

- 22) The course helped me to analyze, interpret and synthesize material.
- 23) This course helped me to reason better and to think more clearly about its subject matter.
- 24) This course helped me to consider alternative perspectives on complex issues.
- 25) I find that this class stimulates my interest in working on this subject outside of class.

Calkins and Micari (2001) suggest that course evaluations are better measures of teaching skill when students are “asked questions that probe learning (such as whether their attitudes or beliefs were changed, whether they understand connections more fully, whether they feel more confident in their ability to tackle problems of the field)” (p. 18). **We recommend that items that probe learning be included in the list of optional items** available to faculty and departments (e.g., see the IDEA Student Ratings of Instruction system at www.theideacenter.org/node/5)

Recognizing the Unique Needs of Online Courses

Table 3 lists items for possible inclusion in an evaluation instrument for online courses. Since an online course often involves posting written work (e.g., essays and student comments) in response to reading assignments and classmate postings, instructor feedback about students'

written work is likely to be seen by students as important. So too is instructor clarity about the nature of the assignments/course design, as well as the students' feelings that they are learning something from reading the publicly posted work of classmates. These issues are reflected in questions 1, 2, 3, and 4 below. Such items would necessarily be course specific and not simply applied to online courses en masse.

Inasmuch as online courses are built around students reading articles and classmate postings, questions that relate to the amount of required reading would also seem to be pertinent. In addition, students may differ in the degree to which they feel they have enough time to assimilate course content. This is because online courses vary in length (online courses can be four, six or twelve weeks long). With these issues in mind, see questions 5, 6, and 7. Again, the questions are course dependent.

Since the instructor is not in a classroom, it is important for him/her to establish a sense of "presence" in the course, as well as to respond quickly to student questions. In that regard, see questions 8, 9, and 10 below. Finally, instructors in online courses find themselves confronted with the task of ensuring that students, while free to express their opinions, do so in a civil manner. Question 11 speaks to this matter. Each item assumes a five point Likert-type scale.

Table 3.

Potential Optional Items for Online Courses

- 1) The professor provided feedback about course assignments in a timely manner.
- 2) I was satisfied with the amount of information the professor provided about the organization of the course.
- 3) Assignment instructions were clear.
- 4) My classmates' postings were educational and informative.
- 5) I was satisfied with the amount of reading the professor assigned.
- 6) Please circle the number that best describes how many assignments the professor required you to complete.
- 7) Please circle the number that best describes your feelings about the length (i.e., number of weeks) of the course.
- 8) The professor promptly responded to any questions I had.
- 9) The professor adequately participated in our online discussions.
- 10) The professor adequately communicated with the class as a whole.
- 11) The online postings of my classmates contained language that offended me.

The Committee recommends that online courses, and perhaps discussion, laboratory, and other instructional settings that necessitate special or different learning environments from

traditional lecture and seminars, such as some arts courses, be assessed through departmental or instructor designed survey items, as described above under Charge 4.

Charge 8: Student Access to SIRF

To address this charge, the Committee examined current practices at other institutions regarding the sharing of course evaluation results with students, conducted a survey of students over the 2010-11 winter intersession, and reviewed numerous published research papers in the course evaluation literature. The survey of practices at other institutions revealed that a number of institutions already share data, with approaches ranging from sharing all quantitative course evaluation data with students and the general public, to sharing only data on specific questions of interest to students for the purposes of course selection. A recent survey by Binghamton University of 90 Chief Academic Officers, Presidents and Chancellors at largely doctoral granting, colleges and universities showed that, of the 41 institutions replying, nine (22%) indicated that they shared course evaluation results with students or external constituents. Our review revealed that five of UAlbany's official peer institutions⁵ share course evaluation results with their students. A Web search of other prominent public research universities identified three additional universities that share evaluation results with students.⁶ The Committee found the University of Colorado-Boulder's [Faculty Course Questionnaire \(FCQ\) Results](#) website to be the most comprehensive and well-constructed of those it reviewed. Boulder's Web site allows students and external visitors to search (or select out) multiple semesters of complete evaluation results by semester, discipline, course number, subject area, instructor group (e.g., tenure track, adjunct, teaching assistant, etc.), and even by individual instructor. The result set can even be exported into an Excel file for customized analyses.

The scholarly literature and empirical data reflects only tangentially on our concerns, and is not plentiful: we found fewer than ten articles focusing on aspects or uses of sharing course evaluation results with students. There is some evidence in the research literature that sharing evaluation results increases student interest in and engagement with the course evaluation process, and thereby increases the likelihood that they will participate (Moss & Hendry, 2002; Nevo, McLean, & Nevo, 2010). Yet we found no evidence at all in the literature that sharing evaluation results leads to less student interest in filling out course/instructor evaluations. There is some dated evidence in the research literature that students use rating information to help inform course selection, primarily based on the perceived quality of the instructor rather than on degree of difficulty or the grading reputation of the instructor (Leventhal, Philip, Abrami & Peery, 1975). This finding resonates with the results of our student survey, discussed below.

⁵ Current Peers: the Georgia Institute of Technology, Binghamton University, and the University of Colorado at Boulder share evaluation results with students. Not sharing results are: Northern Illinois University, Old Dominion University, the University of Connecticut, the University of Hawaii at Manoa, the University of Houston at University Park, the University of Vermont, and the University of Wisconsin at Milwaukee. Aspirational peers: the University of California at San Diego and the University of Virginia share evaluation results with students. Not sharing results with students are: the University at Buffalo; Stony Brook University; the University of California at Irvine, Santa Barbara, and Santa Cruz; and the University of Oregon. UAlbany's peer institutions are defined in the UAlbany – SUNY MOU II, 2006.

⁶ Northwestern University, the University of Maryland at College Park, and the University of Maryland Baltimore County share evaluation results with students.

Based upon the Committee's review of the course evaluation literature and observed practices at other public research universities in sharing evaluation results with students, the Committee decided to ask UAlbany students what information they most wanted from the course evaluations. Guided by practices associated with effective teaching cited in the research literature, the Committee developed an online survey to elicit this information. Students rated a number of items organized around six constructs associated with effective teaching. The organizing constructs are:

- COURSE expectations and requirements;
- COURSE workload;
- COURSE overall rating;
- Quality of INSTRUCTOR-student interaction;
- Effective Student Engagement; and
- Presentation of Course Content

Students rated each item based on how much they would like to see information on these items, ranging from 1 ("not at all") to 5 ("I really want this information!"). **Table 4**, ordered by items students would most like access to, across both undergraduate and graduate students, **shows that students would welcome reviewing evaluation results around most all of the items put forward**. Topline results summaries of the intersession survey are available on the Committee's wiki resource page.⁷

⁷ See the **Student Course Evaluation Feedback Survey** section (in lower right-hand corner) of the Committee's wiki which is found at <https://wiki.albany.edu/display/irpe/Course+Assessment+Advisory+Committee>

Table 4.

Means and Standard Deviations for Items on Student Survey of Interest in Access to Course Evaluation Information

	Mean	N	Std. Deviation
q2.c2 The instructor explained complex material clearly.	4.41	1167	.721
q1.a4 The instructor did a good job of covering the course objectives and content.	4.40	1282	.731
q2.c1 The instructor was effective in communicating the content of the course.	4.37	1172	.713
q1.a1 Course guidelines were clearly described in the syllabus.	4.36	1283	.774
q1.a Course expectations and requirements overall.	4.35	1286	.738
q1.a3 The degree to which examinations and quizzes (or other evaluations) covered the course content and objectives.	4.35	1287	.756
q2.a4 The instructor was approachable and willing to assist individual students.	4.31	1174	.740
q1.b Course workload, overall.	4.31	1288	.805
q1.c Course overall	4.31	1264	.737
q1.a2 The course seemed well planned and organized.	4.31	1281	.720
q2.c3 The instructor used teaching techniques and assignments that promoted my learning of course content.	4.28	1171	.770
q1.b3 The number of course assignments (or projects or papers) was appropriate.	4.27	1287	.819
q2.a2 The instructor was responsive to student concerns.	4.26	1172	.747
q1.b4 Class workload, given the course level and number of credits.	4.26	1274	.833
q2.a1 The instructor's availability for course-related assistance such as e-mail, office hours, individual appointments, phone contact, etc.	4.25	1179	.760
q2.c Presentation of course content overall.	4.23	1178	.767
q2.b2 The instructor helped create an atmosphere that kept me engaged in course content.	4.22	1176	.800
q2.a Instructor-student interaction, overall.	4.15	1179	.792
q2.b1 The instructor's effectiveness in encouraging interest in the subject.	4.13	1179	.814
q2.b3 The instructor challenged you intellectually and held you to high standards of performance.	4.11	1173	.829
q2.a3 The instructor encouraged students to consult with him or her.	4.11	1176	.826
q1.b2 The intellectual challenge of the course.	4.08	1278	.837
q1.c1 Would you recommend this course to a friend?	4.08	1283	.900
q1.b5 Degree of effort you put into the course, given the course level and number of credits.	4.06	1278	.892
q2.a5 The instructor's respect for and professional treatment of all students regardless of race, color, national origin, sex, age, disability, creed, religion, sexual orientation, veteran status, or other personal characteristics.	4.05	1166	.973
q2.b Effective Student Engagement overall.	3.99	1178	.844
q1.b1 The average number of hours per week spent on this course for all course-related work.	3.93	1282	.977

There is growing interest (for example on the part of the Student Association) in accessing course evaluation results. Moreover, independent sources of such information exist, including ratemyprofessors.com and for-hire services. Thus, the committee believes that it is advisable to work internally within the campus community to develop policies and practices for sharing data that accord with the University's values and goals.

Based on our findings from the higher education literature, other institutional sharing of evaluation results with students, UAlbany student opinion, the need to comply with Freedom of Information Law requests,⁸ should they occur, and our own deliberations after considering input

⁸ For example, a 2004 FOIL request regarding course evaluations was satisfied by SUNY Binghamton after the eventual involvement of the State of New York Department of State Committee on Open Government. See

from students and faculty during this report's comment period, ***the Committee recommends that the University make available evaluation results for the course and instructor overall ratings. Specifically, we recommend that the mean overall course and instructor ratings, Ns, standard deviations, class enrollments, response Ns, and percent responding be released.*** While the virtues and detriments of posting students' narrative comments were deliberated, we do not at this time support the public posting of student narrative comments. We recommend that the standing CAAC take up the issue of whether and how to share narrative comments in the future.

In order to encourage students to fill out the online course evaluations, and to garner a high enough response rate for faculty and the administration to be comfortable in using the results for both formative and summative purposes, ***the Committee recommends that, for at least the first year, the evaluation results specified above should only be made publicly available if at least 60% of the students in each class respond to the evaluation survey.*** A 60% response rate is considerably higher than current online response rates, and represents an ambitious goal. We believe it can be met and even exceeded, however, as students become increasingly aware that they need to participate in order to ensure that course evaluation results reflect the views of a substantial proportion of any class's students. The University of North Carolina has a 65% threshold, and the Johns Hopkins University utilizes a 70% threshold. The Committee recommends that the initial 60% threshold be revisited two years after its implementation in order to determine whether or not it can be raised.

The Committee proposes that posting of SIRF results for the course and instructor overall items be phased in by announcing to students and instructors in the fall 2012 semester that the new policy will take effect in spring 2013, at the earliest. Results will be posted for each class with 60% or greater student response rate, and posted results will remain accessible into the future.

Further regarding the public release of SIRF results, ***the Committee recommends that instructors be afforded the opportunity to attach their own comments next to any publicly released SIRF results, should they feel the need to explain or comment on their ratings.*** We recognize that the ability to attach instructor comments to SIRF results might not be technologically possible at this time but recommend that the ability to do so be explored.

Charge 9: Gather and Review Reactions of Students and Faculty

To address this charge item, the Committee invited the comments and reactions of the UAlbany community to its draft report. We are most grateful to the many individuals, and various faculty, governance, and administrative bodies who were kind enough to share their reactions and their constructive criticisms with us. In addition to an open forum, members of the Committee met with the University Planning and Policy Council (UPPC) and the Council on Academic

<http://www2.bupipedream.com/sa-pushes-for-release-of-soot-surveys-1.1807162>. Binghamton subsequently developed a policy to make course evaluation data public.

Assessment (CAA), and we received a written reaction from the Council on Promotions and Continuing Appointments (CPCA). In addition, Committee members met with the College of Arts and Sciences Chairs' Council and with the College of Arts and Sciences Faculty Council, as well as with the Provost's Executive Advisory Committee (PEAC).

In addition to explaining the Committee's thinking or commenting on issues at the time of their communication or presentation to us, the Committee also cataloged the comments and suggestions it received. We then both individually and collectively revisited them. While many of the comments referred to areas in our draft report that were already sufficiently explained, several led us to rethink our initial position. In others, we saw a need to add clarifying language to better communicate our thinking or to further explain the rationale for a particular recommendation.

Summary of Recommendations

1. **The University should periodically reinforce its long-standing guidance in support of a multi-method approach to the assessment of teaching effectiveness.** This recommendation lies at the heart of nurturing a long-term commitment to fostering instructional excellence. The evaluation of teaching should include more than course evaluations, and should be contextualized by data that presents a complete and balanced picture of any individual's teaching practices. The evaluation of teaching should also be forward-looking, by promoting the goals of departments and of their disciplines specifically - and of the university generally - by applying criteria and standards benchmarked to research-based, effective teaching practices. The process of evaluation of teaching will need to be flexible enough to respect both the variety and complexity of effective teaching, while respecting and valuing individual narratives of innovation and improvement. Short of these goals and the mechanisms to support them, a system of evaluation will devolve by default to the simplest and easiest practice, as is exemplified nationwide by the widespread, often exclusive, use of course evaluations in the evaluation of teaching.
2. **A new teaching evaluation instrument and program, which reflects up-to-date teaching assessment methodologies, should be used for multiple functions, such as:**
 - a. Providing consistent information on teaching in support of personnel recommendations for full-time and part-time faculty;
 - b. Improving teaching practice and professional development of full-time and part-time faculty; and
 - c. Meeting student desire for additional information about courses to aid in course selection.
3. **The global items (i.e., instructor overall; course overall) used in university-wide summative course evaluations should be held to a small number of very general items—currently two.** At the college or departmental level, more specific items related to instructor behaviors and to the specific nature of student experience can be relevant for summative purposes if they reflect the values and interests that those units wish to

encourage among instructors. We leave these up to the discretion of the schools/colleges and their faculty.

4. **University-wide results summaries should not include the current more specific analytical questions (i.e., items 1 through 7).** The Committee's analyses have found these items to largely represent overall student satisfaction with the instructor. Additionally, their applicability across instructional settings is highly variable, making aggregate summaries at the school/college or university level inappropriate. Analytic items should be instructor (or course) specific, and should be used by the instructor and/or department for formative assessment purposes.
5. **The University should continue to make instructors' individual and comparative SIRF results available to them as early as possible so that they may use the results to inform their next semester's teaching assignment(s).**
6. **A revised SIRF system should be similar to the one employed by the University of North Carolina, in which assessment items are matched to one of four user groups for purposes of functionality and levels of access:** (1) Institutional research purposes, university-level promotion committees, and other system administrators, (2) deans or department chairs, (3) instructors, and (4) students.
7. **The new SIRF would contain:**
 - a. **Two global questions for students to rate their overall impression of the instructor and course.** These items would be used for summative evaluation of the instructor by departments, deans, and the Council on Promotions and Continuing Appointments (CPCA). Summary results would also be provided to students to assist in course selection.
 - b. **Two or three departmental questions which are multiple choice or Likert-type survey items, and/or open-ended text questions that the department sponsoring the course wishes included for purposes of formative assessment.** These items could be developed internally by the department, or selected from items like those already in use at other institutions (e.g., the UNC item bank). These items may or may not be part of faculty tenure and promotion portfolios.
 - c. **Two-to-five instructor-designated multiple choice, Likert-type, and/or open-text items.** Item summaries would be restricted to the sole province of the instructor.
 - d. **Expand the Likert-type scale used on the survey instrument to seven categories from the five currently in use,** as the Committee believes this would capture more variability in the student response set, and would therefore be more useful to faculty.

8. **Students should be asked to select the instructor's name from a dropdown menu in each evaluation as a reminder of which instructor they are rating at any particular time.**
9. **SIRF results for the overall course and instructor items should only be shared with students for those courses that achieve a student participation rate of 60%.**
Providing access to evaluation results (i.e., mean, standard deviation, N, class enrollment, percent responding) only when a sizable percentage of students in each class complete their surveys seems reasonable, and is likely to increase the response rates for the online SIRFs.
10. **The University should explore the means to allow instructors to include their own comments next to any of their publicly posted SIRF results, should they desire.**
11. **Instructors should communicate to students how they utilize SIRF results, both the global items, and the instructor and departmentally designed items. Furthermore, each academic department should reinforce this communication stream with its tenure-track and adjunct faculty within its teaching policies.**
12. **Online and arts courses which are not lecture style, and perhaps discussion, laboratory, and other instructional settings that necessitate special or different learning environments from traditional lecture and seminars should be assessed through additional, distinct and separate departmental or instructor designed items.**
13. **A standing Course Assessment Advisory Committee, under the auspices of the Provost, should be established, with rotating membership from among the deans, department chairs, faculty governance, student leaders, and others involved in personnel decisions.** This standing committee should meet at least once in the beginning of each semester to review, advise, and recommend changes to course assessment procedures and/or policies. This will give the faculty continuing input on the implementation of changes and opportunities to set policy as questions arise. While not discussed in the report, establishing a standing committee to fulfill these functions and provide ongoing, rather than episodic, input into course assessment policy formulation seems sensible.

Issues for the Standing Course Assessment Advisory Committee

This report referenced a number of issues that should be addressed by a new, standing Course Assessment Advisory Committee in the future. These issues include:

- The development of tools and techniques for assessing teaching, in addition to the SIRF, as part of the multi-method approach to which the University subscribes. For example, the report recommended developing and publishing examples of possible evidence for the evaluation of teaching that could be included in a teaching dossier, including but not limited to narratives, portfolios, and peer reviews.

- The need to build through policy an expectation and a support process for regular formative assessment of teaching for faculty, with a shorter frequency (1- 2 years) for pre-tenure faculty.
- The need to develop tools or processes to guide administrators' evaluation of a faculty member's teaching.
- Making explicit the expectations for effective teaching, and sharing with new hires the written standards and instruments for evaluation of teaching.
- The effectiveness of the new SIRF instrument and the revised summary report in serving both formative and summative purposes for all types of courses.
- The effectiveness of the 60% response rate threshold for sharing SIRF results with students in raising response rates.
- The development of guidelines for the formatting of instructor and departmental survey items (scale construction, item stem and response categorization, etc.) and what the approval process will be, if any. For example, are items to be chosen from an approved list, or should the University allow latitude to instructors and departments? While the report recommended some latitude, some type of centralized oversight seems prudent.
- The need for a SIRF program administration protocol that instructors and departments follow in order to inform students of the purposes of the SIRF and the ways in which SIRF results are put to use.
- The establishment of parallel protocol standards for departments that wish to develop and administer their own course/instructor evaluation forms.
- The usefulness of sharing narrative course evaluation comments with students. The 2012 report did not recommend sharing narrative comments because of technical difficulties with managing them (e.g., redacting inappropriate comments, concern that outliers would garner most of the attention, etc.) but the Committee recognized the value of the comments in interpreting the numerical ratings of instructors and courses.
- The feasibility of asking students to use personal electronic devices such as smart phones, laptops, or tablets to complete the SIRFs during class time.

References

- Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., Bell, D. (2006) Electronic course evaluations: Does an online delivery system influence student evaluations?" *Journal of Economic Education*, 37(1), 21-37.
- Calkins, S., & Micari, M. (2010). "Less-Than-Perfect Judges": Forty years of discourse on the value of students rating teachers in higher education. *Thought and Action*, 7–22.
- Cizek, G. (2010). An introduction to formative assessment: History, characteristics, and challenges. In Andrade, H. & Cizek, G. (Eds.) (2010). *Handbook of formative assessment*. New York: Routledge.
- Cohen, P.A. (1981). Student ratings of instruction and student achievement: a meta-analysis of multi-section validity studies. *Review of Educational Research*, 51(3), 281-309.
- Dommeyer, C. J., Baum, P., Hanna, R. W., Chapman, K. S. (2004) Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations." *Assessment & Evaluation in Higher Education*, 29(5), 611-623.
- Feldman, K. A. (1998). Identifying exemplary teachers and teaching: evidence from student ratings. In K.A. Feldman & M.B. Paulsen (Eds.), *Teaching and learning in the college classroom*. Needham Heights, MA: Simon & Schuster.
- ITLAL Peer Observation and Assessment of Teaching (2008). Available at http://www.albany.edu/teachingandlearning/tlr/peer_obs/Peer%20Observation%20Resource%20Book%20for%20UAlbany.pdf
- Leventhal, L., Abrami, P., Perry, R., & Breen, L. (1975). Section selection in multi-section courses: Implications for the validation and use of teacher rating forms. *Educational and Psychological Measurement*, 35, 885-895.
- Moss, J., & Hendry, G. (2002). Use of electronic surveys in course evaluation. *British Journal of Educational Technology*, 33(5), 583-592.
- Murphy, S.E. (1987). Mirroring instructional techniques: A process evaluation model. *Training and Development Journal*, 104-107.
- Nevo, D., McClean, R., & Nevo, S. (2010). Harnessing information technology to improve the process of students' evaluations of teaching: An exploration of students' critical success factors of online evaluations. *Journal of Information Systems Education*, 21(1) 99-109.
- Office of Academic Affairs (2005-06). Appendix H: Senate Bill 8384-07 on Evaluation of Teaching. Available at: http://www.albany.edu/academic_affairs/policies_guidelines/continuing.shtml
- Office of Institutional Research, Report of the Pilot Test of the Student Instructional Rating Form (SIRF). (1982). Available at: https://wiki.albany.edu/download/attachments/19674280/Full_Pilot_Test_Report_1982.pdf?version=1&modificationDate=1269480751000
- Pajares, F., Hartley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development*, 33(4), 214-221.

Seldin, P. (1991). *The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions*. Bolton, MA: Anker Publishing Co.

Urbach, F. (1992). Developing a teaching portfolio, *College Teaching*, 40(2), 71-74.

Members of the Course Assessment Advisory Committee

- Heidi Andrade (co-chair) - Associate Professor, Educational Psychology and Methodology; Past Chair, Council on Academic Assessment
- Bruce Szelest (co-chair) – Associate Vice Provost for Academic and Resource Planning
- Laura Benson-Marotta - Research Analyst, Institutional Research, Planning, and Effectiveness
- Indushobha Chengalur-Smith - Associate Professor and Chair, Information Technology Management
- John Delano - Past Chair, UPPC; Distinguished Teaching Professor, Atmospheric and Environmental Sciences
- Jamie Fader - Assistant Professor, Criminal Justice
- Kevin Fallon – Undergraduate Student
- Richard Fogarty - Associate Professor, History
- Richard Hamm - Professor and Chair, History
- William Husson - Visiting Assistant Professor, Communication
- Joseph McAneney – Undergraduate Student
- Olivia Napoleon – Undergraduate Student
- Saggi Nevo - Assistant Professor, Information Technology Management
- Julie Novkov - Professor, Political Science, CPCA member
- William Roberson - Director, Institute for Teaching, Learning, and Academic Leadership
- Sylvia Roch - Associate Professor, Psychology
- Leah Rotella - Vice President, Student Association
- Gregory Stevens - Assistant Dean for Academic Programs, College of Arts and Sciences
- Sarah Taylor - Vice President, Graduate Student Organization
- Anna Valcheva – Graduate Student

Appendices A - G

The following appendices are available at <https://wiki.albany.edu/display/irpe/chairs+page>

- Appendix A: Base Statistics and Rating Breakdowns by Various Student, Course, or Instructor Attributes
- Appendix B: Bivariate Correlations and Full Regression Model Results for In-class and Online Evaluations, with Instructor Overall Rating as Criterion Measure
- Appendix C: Bivariate Correlation Results for Each Administration Mode; In-class and Online
- Appendix D: Factor Analyses
- Appendix E: Results of Reliability Tests
- Appendix F: UNC's Functionality Grid
- Appendix G: A Proposed New UAlbany SIRF Form