Academic Affairs 2012 Task Force on a University-Wide System of Student Input on Teaching Effectiveness

Academic Affairs Committee (AAC)
Task Force Members
Roxanne Canosa*
Carol De Filippo*
Liz Kmiecinski
Elizabeth Kronfield, co-chair
Michael Laver, co-chair*
Christine Licata, Provost delegate*
Jeff Lodge
Erhan Mergen
Ed Mineck
Frank Sciremammano
Linda Tolan
Walter Wolf
Tracy Worrell*

Richard Latham, student government representative*

*AAC subcommittee members responsible for preparing this report
Table of Contents

I. EXECUTIVE SUMMARY 3

II. TASK FORCE CHARGE 5

III. BACKGROUND 7
   A. The 2012 Task Force process in addressing its charge
   B. Issues to be addressed: Community response to the 2010 Task Force
      recommendations
   C. Language confusion and need for clarity of focus
   D. A comprehensive approach: Summative and formative functions
   E. Faculty concerns
   F. Considerations for selection of survey items
   G. Best practice in administration of student ratings

IV. RECOMMENDATION FOR A UNIVERSITY-WIDE SYSTEM OF STUDENT INPUT
    ON TEACHING EFFECTIVENESS 19
    A. Overview and rationale
    B. Pilot phase
    C. Considerations for university-wide launch

V. RECOMMENDATIONS FOR SUPPORTING A SYSTEM OF STUDENT INPUT
   ON TEACHING EFFECTIVENESS 27

References 30

Appendix A: Outline of Supplemental Guidelines Appended
            to the Task Force Charge 32

Appendix B: Executive Summary of Written responses to
            the 2010 Task Force Core Items 33

Appendix C: Reflections on Major Research Findings on Student Ratings 38
I. Executive Summary

Charge
The Task Force on Student Input on Teaching Effectiveness was charged this year with reviewing multiple sources of feedback regarding student ratings, establishing a set of principles that would inform the development of an online student rating system, and identifying guidelines by which such a system would be implemented at RIT.

Actions
The Task Force first set about 1) evaluating extensive feedback from RIT faculty and colleges, and collated that feedback. We then 2) performed a comprehensive literature review on the subject of student ratings in general and online student ratings in particular. Next we 3) identified two vendors out of a final list of four that met our needs and that were suitably different so that the RIT community would have a genuine choice as to what model of online system suited RIT. Both systems could accommodate summative and formative items. We then 4) established a set of core questions that would be asked in the customizable survey. And finally, we 5) arrived at the need for a pilot and established some preliminary guidelines as to how the pilot would be implemented in Fall quarter 2012.

Results
1. Our literature review revealed that student rating of instruction is both reliable and valid as an indicator of teaching effectiveness. We learned that students prefer the online method and that generally, instructor ratings do not vary in a statistically significant way in online versus traditional ratings. We learned that students will generally provide much more qualitative data online than they do in traditional surveys. And finally, we learned that the non-response bias that tends to be a factor in faculty resistance is not borne out in the research, although the research does suggest that response rates may be lower in online surveys and that measures may have to be adopted to try to boost the rates of response.

2. The Task Force identified two vendors to pilot: The IDEA Center's IDEA Tool and SmartEvals from GAP Technologies. The IDEA Tool is a comprehensive system that requires faculty to input desired learning outcomes and requires students to input various data, and from that data the system provides each faculty member with a customized result that highlights where faculty members might improve their teaching, complete with references to materials on teaching effectiveness. SmartEvals, on the other hand, is essentially a platform for hosting a survey of our own creation, although as one can see from the full report, there are quite a few features of the technology that are attractive: for example, the ability to dig deeper in a particular response to see how the rater answered.
other questions on the survey in an attempt to contextualize responses. We recommend that these two systems be piloted in fall term 2012.

3. The Task Force developed a set of core items for the SmartEvals survey. The IDEA Center does not require this as they provide their own items. We have used best practices as indicated in the literature to develop a single global summative item as well as a few formative items to be used across the University. In both systems, colleges, departments, and individual faculty members can add items of their choosing.

4. The Task Force established rudimentary principles that are meant as a starting point for a body of faculty who will oversee the implementation of the pilot programs next year as well as the collection of feedback to arise from the pilots. We recommend that the Academic Affairs Committee form the core of this group, drawing on expertise from outside as needed.

Conclusion
The Task Force feels that the successful transition to an online system of student input on teaching effectiveness is best accomplished with as much faculty buy-in as possible. To that end we feel that, just as this Task Force was faculty driven, the pilot, the feedback from the pilot, and the implementation and review of the new system be primarily faculty driven. We also feel very strongly that when the new system is implemented, there needs to be put into place a regular system of opportunities for faculty to be able to improve their teaching through programs in the Teaching and Learning Center and possibly through other groups as well. We firmly believe that if the faculty can utilize student ratings of instruction to improve their performance in the classroom, then the system will benefit all involved: students, faculty, and administrators.
II. TASK FORCE CHARGE

The Academic Affairs Committee received a formal charge from the Academic Senate on December 1, 2011. This charge included the following action steps:

1) Identifying a compilation of issues, concerns, and suggested solutions extracted from the feedback received regarding the Course Evaluation Task Force Report of August 15, 2010. Attention should also be paid to the feedback from the colleges, and in particular, from NTID.

2) Establishing principles that would govern development of a university-wide student survey.

3) Establishing guidelines for administering a university-wide student survey, including purpose, scope, and timelines for a phased implementation, that issue from the established principles and follow best practices as summarized in the literature.

4) Drawing up a revised set of core questions developed in consultation with on-campus survey experts and consonant with documented best practices.

5) Identifying best practices to be employed in the redesign of the evaluation of teaching in order to be of value to faculty and students.

The taskforce should be comprised of a representative from each college and a representative from student government. An ex officio member will be appointed by the Provost from Academic Affairs and an additional ex-officio member from ETC in the Wallace Center to address technological issues. The taskforce, in consultation with the AAC, will report back to the senate with its recommendations by the end of academic year 2011-2012.

Background information provided to the Task Force regarding the charge

Over at least the last decade, various plans have been considered to develop a university-wide process to allow for student input as one component of annual review of faculty, per policy 7.B.5.c:

"B.5. Each college's or academic unit's review process must include the following elements: … Standardized student evaluations as established by college policy."

In an attempt to achieve multiple objectives, these plans have met with considerable controversy. Given the immediate need for new technology to administer any student survey process, crafting a plan that satisfactorily addresses faculty and administrators' concerns and follows best practices in this area has now become a pressing issue.

Based on the strength and substance of feedback to the most recent plan described in the Course Evaluation Task Force Report of August 15, 2010, the AAC has determined the need to clearly identify, separate, and prioritize competing objectives and draft a recommendation
that addresses a phased approach to survey development and administration, including specific indicators of readiness to move from one phase to the next. The recommended plan will resolve the distinctions between:

a) Review of faculty vs review of courses (established curriculum evaluation processes)
b) Student perceptions of their teachers' instruction vs administrative evaluation for annual review
c) Need for formative assessments to improve teaching effectiveness vs administrative needs for summative data for cross-university comparisons
d) Confidential matters of personnel deliberations vs students' desire for access to their peers' opinions

The plan will also be responsive to the following needs:

e) Input/collaboration during the development process involving faculty, on-campus survey experts, and students
f) Education and communication with/for the community regarding best practices, addressing relevant distinctions (per the list above), and clarifying the desired scope and objectives of the plan
g) Periodic assessments as phases of the plan are implemented, including summaries of survey results, and their implications for faculty
h) Criteria for justifying a widening of access to survey results in future survey administrations, if found to be warranted

Supplemental guidelines based on the NTID feedback were also provided. Those recommendations appear in Appendix A.
III. BACKGROUND

"Viewing student ratings as data rather than as evaluations may also help to put them in proper perspective."
Cashin, 1995

"We should not confuse a source of data with the evaluators who use it – in combination with other kinds of information – to make judgments about an instructor’s teaching effectiveness [Cashin, 2003]."
Benton, 2012, p. 13

Research and best practice have not changed very much over the years with respect to student input on teaching. It continues to be widely recognized and confirmed by research that students can provide important data to be used for an administrative evaluation of teaching. Our recommendations align with this very context, which Provost Haefner affirmed in his charge to the 2010 Task Force on this issue. That is to say, student input is but one of a variety of types of information necessary to render an evaluative judgment about teaching effectiveness and, ultimately, overall faculty performance.

A. The 2012 Task Force process in addressing its charge

The 2011-2012 Academic Affairs Committee (AAC) of Academic Senate (AS), serving as the Task Force charged by the AS, designated a subgroup to spearhead the development of recommendations. The subgroup reported periodically to the full committee, which provided feedback and guidance and approved the final recommendations contained herein.

Work within the subgroup proceeded in several phases:
• Review of research on student ratings
• Examination and analysis of community input on the report of the 2010 Task Force on Student Evaluation of Courses and Instructors (Appendix B contains a summary of themes extracted from that feedback)
• Consultation with the RIT Office of Legal Affairs regarding confidentiality of student ratings
• Review of the recommendations of the 2010 Task Force
• Discussion and delineation of the desired characteristics of a student input system in light of the guidelines in our charge
• Identification of commercial systems that appeared suitable for further investigation
• Interviews with representatives of each of the identified systems
• Discussion of system strengths and weaknesses
• Drafting of recommendations based on the research, community input, guidelines in our charge, and responses from potential vendors
• Discussion of recommendations with the full Task Force
• Confirmation of support from the full Task Force of the final recommendations and report

B. Issues to be addressed: Community response to the 2010 Task Force recommendations

In consideration of the 2010 Task Force recommendations, faculty raised a variety of issues, via anonymous comments on a Clipboard survey (Appendix B) and in an open conversation via email. In addition, detailed feedback from NTID was provided (outline of recommendations appears in Appendix A). There was a sense among many that the recommended core items just did not capture the essence of the respondents' vision for this student input tool. The recommended survey items received considerable critique in terms of focus (whether the items were relevant to the desired focus) and wording. There were also comments about practices surrounding the use of student input and its weighting as part of the annual performance appraisal and tenure and promotion recommendations. An issue that appeared to be particularly unsettling was the Task Force recommendation to provide public access to student input.

Our review of the community feedback, coupled with the very valuable work of the 2010 Task Force and our own investigation into best practices in the field, led us to organize our thinking in terms of four aspects of any system of student ratings of instruction. Important issues associated with each of these aspects, and strategies potentially useful for addressing them, served as the foreground for our ultimate set of recommendations.

1. Purpose
   This issue is the "why" of student input. What does University policy say regarding the focus of and use of student input? Are policy changes needed to better reflect best practice? What is the role of formative versus summative feedback in regard to faculty development, annual appraisal, tenure, and so on?

2. The survey instrument
   What items are appropriate for a university-wide student survey (whether summative or formative)? What are the research findings regarding areas in which students provide reliable and valid input? Can the survey items in the 2010 Task Force report be acceptably modified in line with feedback from the faculty survey and email exchange, responses from the colleges, and published research? Or do these items need to be replaced?

3. Access and interpretation
   What are the issues related to privacy and the legality of publishing a component of a personnel decision (the annual performance appraisal)? The Task Force desired information from RIT legal advisors before making a recommendation on levels of access (such as faculty member only, department chairperson, dean, all students at RIT, the general public). Also, what are the
implications to consider in our recommendations regarding how the results of the student input are displayed or how they might be interpreted by a chairperson or tenure/promotion committee?

4. Administration
What are the needs for educational programs that arise from the administration, interpretation, and application of findings from a university-wide student survey? What issues arise related to the implementation and continuation of University support for the individual(s) charged with implementing, monitoring, and maintaining the system; for department chairpersons in preparing annual appraisals; for faculty who wish to access, customize, administer, or encourage participation in a survey; for faculty who seek assistance in improving aspects of their teaching effectiveness; for student respondents who seek accountability? Other administrative issues pertain to the timing of the survey (when it should be open to students); incentivizing student participation; ensuring an adequate response rate; managing data summaries and reports over time to account for low-enrollment courses; and designating resources to monitor, document, and analyze the impact of student ratings on faculty, students, and teaching across the university over time.

C. Language confusion and need for clarity of purpose

"Teaching effectiveness." "Teaching evaluation." "Course evaluation." "Course and instructor input." How we have labeled the what and the why of a university-wide system of obtaining student input has contributed to paralyzing and polarizing attitudes and confusions about how to go about doing so. Provost Haeffner, in his November 2010 memo to the RIT faculty on a "Student survey for teaching evaluation," established a guiding theme in his statement that: "…we must be able to measure teaching effectiveness…." Thus, in formulating their core items, the 2010 Task Force consulted research findings on the components of teaching effectiveness (e.g., Adams & Pierce, 2004; Berk, 2005; Cashin, 1990, 1995, 1996). Despite this effort, some unfortunate terminology crept into the discussions. The Task Force, itself, was named variously "Task Force on Student Evaluation of Courses and Instructors" and "Course Evaluation Task Force." Its charge referred to both "student evaluation of courses and teaching" and a request to draft "core questions that adequately reflect the essentials of teaching excellence." The Task Force report was entitled, "Course Evaluation Task Force report," but the survey content was "aimed at evaluating the faculty member’s effectiveness as a teacher." The result is a menu of disparate choices.

1. Choice of focus
Evaluation vs measurement. The term evaluation can denote a judgment of value and, in administrative matters, that has been the province of a faculty member's supervisor (per Policy #E7.0). Further, there may be unique weights assigned to a given individual's responsibilities including, not only effective teaching, but also professional and academic qualifications,
professional activities, contributions to the university, and community activities. And, within the area of teaching, "standardized student evaluations" are only one type of documentation to be obtained.

**Teacher vs teaching.** Use of the terms teacher and teaching suggests a focus on either the individual or the art of teaching, itself. Those who study student ratings have concluded that it would be more useful to focus on students' ratings of what they gained relative to the course objectives (McKeachie, 1997) rather than the individual teacher or the students' conception of how to teach ("teaching"). Student ratings are valuable for informing teachers about student expectations, frequency of opportunities provided for active learning, apparent preparedness of the instructor for each class, and methods used for assessing student work (McGowan & Graham, 2009).

**Teaching vs effectiveness vs excellence.** Different kinds of learning often call for different types of teaching (Hoyt & Lee, 2002). Thus, it may not be possible to create a consistent university-wide survey that applies to the varieties of teaching required across the curriculum. In contrast, effectiveness speaks to desired outcomes as measured from such sources as peers, alumni, or exam grades. One of the difficulties is that student learning is not a function merely of one's teaching effectiveness; rather, certain student characteristics (such as motivation or ability) are known to have an impact on student learning (Kulik, 2001). Moreover, the typical sources of outcome measurement are associated with complications that are difficult to overcome. Exams do not uniformly exhibit adequately high reliability. Useful peer observation is dependent on training. And alumni feedback does not contribute new information over and above what can be learned from currently enrolled students (Benton, 2012). Nonetheless, it is teaching effectiveness that has been the subject of most research on student ratings and, although "Rochester Institute of Technology is committed to promoting academic excellence" as an achievement to strive for, it is effective teaching that is the subject of the annual performance review at RIT (per policy E7.B.2) and about which students are qualified to comment.

2. **Defining the effective teacher**

As noted by the 2010 Task Force, the effective teacher is described in RIT policy in relation to the criteria for granting tenure. According to this policy, the effective teacher has special knowledge and expertise, teaches with sensitivity toward students, and applies appropriate teaching skills:

E5.0 Policy on Tenure
2. Conditions of Tenure Appointments
d. Criteria for Granting Tenure
(5) An effective teacher, among other things, communicates special knowledge and expertise with sensitivity towards students’ needs and abilities. This entails selection and use of appropriate instructional methods and materials and providing fair, useful, and timely evaluation of the quality of the learner’s work.
Many faculty who responded to the 2010 Task Force report supported a focus on teaching effectiveness, but there were varying expectations that items would measure "excellence in education" in a broad sense; would allow a measure of "how students will apply their learning" in the future; would be "linked to an assessment of learning outcomes, as related to the course objectives -- quantity and quality of learning achieved;" or would assess specific teacher attributes. Faculty expressed widely ranging opinions as to how satisfactorily the proposed survey items addressed relevant teacher characteristics.

**D. A comprehensive approach: Summative and formative functions**

According to Cashin (1996), clarity as to why we collect student data about faculty is second in importance only to defining institutional mission. He underscored the clear distinction between the summative functions of student surveys (for "evaluation-making personnel decisions") and the formative functions (for "faculty improvement"). Cashin also cautioned strongly that the uses of faculty evaluation data must be decided before any data are collected as a matter of fairness and defensibility and to increase the likelihood of acceptance by faculty. He advised that students join faculty and administrators in a discussion of the purpose for collecting student input on their instructors and cautioned that, "not all student rating items serve every purpose equally well" (Cashin, 1990).

1. **Faculty improvement through student ratings**
   For improvement, Cashin (1990) recommended a student rating system that is flexible, as well as diagnostic. A flexible survey recognizes that effective teaching methods will differ from course to course to the extent that course goals differ. Thus, items are to be selected by the instructor based on course and student characteristics. A diagnostic survey contains items that are specific and behavior-focused so that they can be useful for improvement. RIT Policy E7.A is consistent with the idea of a flexible diagnostic survey: "An underlying principle of this policy is that faculty evaluation and faculty development are closely related and work in concert to help faculty meet individual and institutional goals."

   Faculty commenting on the 2010 Task Force recommendations also expressed a need for a dual-purpose survey to "get teaching of the highest quality" by "helping a teacher to improve." As one respondent put it, the student survey should help to create a "culture of effective teaching."

2. **Faculty evaluation through student ratings -- one source of evidence**
   In Benton's (2012) recent review of the literature on student input, summary evidence is provided to document that student input is valid and reliable when focused on certain dimensions of instruction, and is unrelated to teaching effectiveness when focused on other dimensions. For example, students cannot validly judge such aspects as whether the objectives for a course are appropriate, if the assignments or readings are the most relevant, or to what extent the content is
current. Also, validity will be affected by certain sources of student bias (e.g., student motivation and eagerness to learn; general overall impression of the teacher) and statistical relationships among course, instructor, and student variables that are potentially operating during survey development, data summary, data reporting (statistical analyses of the ratings), and data interpretation. Some of the variables that are related to student ratings, and which call for use of comparative data or score adjustments are course level, class size, academic discipline, course workload and difficulty. Validity is also compromised if student anonymity is not guaranteed, if the teacher is present during survey administration, and if student perceptions about the uses for the survey results are not controlled.

These findings about the many sources of bias and effects suggest the following considerations:

• Invest in a survey that incorporates the necessary analyses to account for known effects
• Implement an education plan for students regarding the purposes and uses of survey results
• Use an online administration to protect student anonymity
• Emphasize that ratings from students must be viewed in light of additional evidence of teaching effectiveness, which often includes such sources as: alumni ratings, peer ratings, personal self-assessment statements, syllabi and other course documents, evidence of student work and progress, and teaching portfolios

3. Policy statements on faculty evaluation and faculty development
University policy E7.0 on Annual Review of Faculty provides a context for obtaining student input. It is noted that the policy begins with an emphasis on using the annual review to foster ongoing professional development and improvement, as well as to make employment and salary decisions. The policy reads in part:

E7.0
A. Preamble
… The results of the review will be used to:
1. Encourage and foster continued professional development.
2. Provide part of the required documentation in the pre-tenure review process of tenure-track faculty.
3. Promote the improvement of individual performance.
4. Determine annual merit increments.

An underlying principle of this policy is that faculty evaluation and faculty development are closely related and work in concert to help faculty meet individual and institutional goals.

The policy continues with recognition of the breadth of considerations that must enter into employment decisions, only one of which is teaching:
E7.0

B. Review Process
2. The criteria for the review shall be consistent with the performance criteria in the university policies for tenure (E5.0) and promotion (E6.0); these criteria include effective teaching, professional and academic qualifications, professional activities, contributions to the university, and community activities.

5. Each college's or academic unit's review process must include the following elements:
a. The faculty member's written self-evaluation and evidence of performance in the criteria specified above (B.2.).
b. As part of the self-evaluation, the faculty member's plan of work for the following year, and discussion on achievement of goals established in previous plans of work. …
c. Standardized student evaluations as established by college policy.

While students are likely unaware of their instructors' content training, curriculum development expertise, materials preparation, or delivery of other courses in which they were not enrolled, they can provide quantifiable observations of their experiences. Students can offer their reflective opinions, comments, and appraisal (i.e., ratings) of instructors in the context of a specific course. Student input then becomes one of several measures contributing to an overall administrative evaluation for personnel decisions such as tenure review and annual merit increments.

E. Faculty concerns

1. Relevance of student input
Deep concerns were expressed by faculty about the legitimacy of student input, with intimations that students are unable to comment in a relevant manner regarding the value of their instruction. For these faculty members, students are supplying worthless opinions or are unwitting participants in a marketing ploy. In particular, some faculty do not see the student survey as a valid basis for guiding efforts to improve instruction. Rather, it appears to be "just a student opinion scale rather than an evaluation instrument" that "measures present satisfaction that may not be related to the deep efficacy of learning." It "appears to promote 'consumerism' in education by giving students what they want rather than what they need." In stating that student input is not relevant to the goal of teaching effectiveness, the view was expressed that, if we want to improve education of students, then we should look at the "learning process – curriculum, requirements, etc." (i.e., a course/program focus).

Other faculty remarked that accountability is a rightful purpose of surveying students. These individuals agreed that the benefit of a student survey is "to give students a voice regarding quality of their classroom experience – to have impact on instruction" and "for student reflection on what the student got out of a course." Faculty noted that seeking student input was an affirmation of the students' role and contributed to their development as learners. For these
faculty members, the student survey is about "being accountable to our students, our curriculum, and our profession."

2. Negative ramifications
Faculty commented on the undue weight that student input has received and expressed grave concern that the proposed survey will serve a punitive purpose. Faculty reported that, in their experience, "annual assessments have relied 100% on the student ratings." Concern arises from observations that "tenure, promotion, and salary decisions are influenced by data from it" and "there is a lot at stake for us." Student survey findings "are a way to intimidate faculty." The "premise behind evaluations is that there is something wrong with what we do." Faculty feel that interpretation of student input "should be done in partnership with the faculty member," but that that partnership has not been realized. In consideration of these concerns, a number of faculty members concluded, "There is no need for public access!"

3. Questions of validity and reliability
It has been argued that students have access to limited evidence regarding the value of a faculty member to the university and that they are capable of commenting on only selected facets of effective teaching. In his recent summary of research on student ratings, Benton (2012, p. 2) summarized his findings on how student ratings are viewed and why:

   The following are some of the most commonly held misconceptions:
   - Students cannot make consistent judgments.
   - Student ratings are just popularity contests.
   - Student ratings are unreliable and invalid.
   - The time of day the course is offered affects ratings.
   - Students will not appreciate good teaching until they are out of college a few years.
   - Students just want easy courses.
   - Student feedback cannot be used to help improve instruction.
   - Emphasis on student ratings has led to grade inflation.

   These myths ignore more than 50 years of credible research on the validity and reliability of student ratings. They persist, unfortunately, largely due to ignorance of the research, personal biases, suspicion, fear, and general hostility toward any evaluation process (Theall & Feldman, 2007).

Appendix C contains a reflection on some of the major research findings on student ratings of teaching effectiveness. Based on a consideration of the preponderance of the evidence reviewed, it was the opinion of the Task Force that student survey responses can meet a sufficiently high level of validity and reliability so as to serve as one component of annual review when constructed and administered within the procedural guidelines established by research.
F. Considerations for selection of survey items

Given the importance that RIT places on teaching effectiveness and establishing a supportive learning environment for students, the Task Force considered developing 3-5 items that would apply across the university, to be seen as formative in nature. In addition, consistent with research in the area of student ratings, we considered 2-3 global items with a summative function. These selective formative and summative items would limit the length of the university-wide survey, allowing for the flexibility to add items of unique interest to individual faculty, departments, and colleges.

For evaluation purposes, a few global (summary) items have been recommended by researchers because they yield generally higher correlations with student learning compared to specific items (Cohen, 1981). A short list of summary items suggested by Cashin (1990) includes these three:

1. Overall, how effective was the instructor?
2. Overall, how worthwhile was the course?
3. Overall, how much did you learn?

Such summary items are analogous to a course grade given to a student in that both communicate information about the level of performance in general but without specifics. More important than the number of survey items, however, is adherence to the principle that generalization of the survey results to assessment of an instructor's overall teaching effectiveness requires a broad sample of courses and terms (Cashin, 1990).

In his recent review of student ratings research, Benton (2012), agreeing with McKeachie, favored connecting student ratings with specific course goals and objectives:

"Offering a relevant view, McKeachie (1997) argued that, when it comes to personnel decisions, student ratings of attainment of educational goals and objectives are preferable to multiple dimensions or a single measure of overall teaching effectiveness. Effective teaching can be demonstrated in many ways, and no instructor should be expected to demonstrate proficiency in all methods and styles. Moreover, teaching methods may vary, depending upon the course content, student characteristics, and size of class."

G. Best practice in administration of student ratings

1. The mechanics
In terms of the mechanics of survey administration, one of the major faculty concerns with an online system of student input remains participation rate. There is a fear, when there are low numbers of responses, that outliers will have an undue influence on ratings. Research findings have allayed these concerns, indicating that averaged ratings are not negatively affected by frivolous responses or by purportedly disgruntled students. While paper surveys typically capture more students, students seem to prefer an online system (Dommeyer, Baum, & Hanna,
2002) and tend to provide more text responses online (Anderson, Cain, & Bird, 2005; Donovan, Mader, & Shinsky, 2006).

Adequate response numbers remains an issue of concern for online systems, but it is also one that is readily manageable, potentially yielding return rates that are equivalent to or better than paper surveys (Wode & Keiser, 2011). The IDEA Center (2008) compiled a list of strategies that universities have found to increase online responses:

- Communication from instructors to students, encouraging them to provide feedback
- Additional student communications (ad campaigns) such as ads, posters, radio announcements, and mailbox flyers
- Reassuring student that their responses are confidential
- System monitoring by an on-campus coordinator and notification to faculty about the status of response rates during the survey period
- Acknowledging or rewarding individuals or units with good response rates
- Using class time where and when feasible to complete the survey
- Verifying email and login accuracy to ensure student access
- Frequent reminders during the survey period and flexibility on end dates if needed to attain adequate response numbers
- Integrating the process into the campus culture, beginning with freshman orientation, emphasizing student responsibility to participate in the improvement of learning

Other sources of concern pertain to the calculations that would enter into a faculty evaluation. It has been recommended to average across surveys from 6-8 classes, representing the variety of teaching responsibilities for the individual, not necessarily conducted during the same academic year, and supplementing these with more classes when there are fewer than 10 students in a class (The IDEA Center, 2004). The index of teaching effectiveness can be criterion-referenced or calculated on a comparative basis. A criterion-referenced system requires that the university establish standards (numerical cutoff points) for each performance category (excellent, good, poor, and so forth). A normative approach judges teaching effectiveness against a comparison group, either within the university or in a national sample, perhaps for a given discipline. In either the criterion-referenced or normative approach, an additional consideration is whether or not the system will adjust for influences that the teacher cannot control, including course difficulty, class size, and student motivation, work habits, and effort. If score adjustments are desired, the system selected for survey administration must allow for collection of relevant student and course information and provide the option for calculating the adjustment.

2. The bigger picture
Cashin (1996) set out 20 principles – based on strong agreement found in the literature – to be applied in developing an effective faculty evaluation system, one that is used to make personnel decisions. The principles are listed here, along with implications for how each is relevant to the Task Force charge.
1. The institution – and the units within the institution – must develop clear goals
   Consult the mission and goals of RIT, as reflected in university policy, as they pertain to
   faculty evaluation.

2. Decide on the purpose(s) data will be used for before any data are collected
   Stipulate that data collection will serve both a summative and a formative purpose.

3. Use pilot programs when appropriate
   Try the system out before committing to significant changes.

4. Significantly involve participants – especially campus leaders – in the development of the
   system
   Cultivate acceptance and ownership by involving all constituencies in system
   development, including faculty, students, and administration.

5. Foster extensive, open communication before, during, and after the adoption of the
   system
   Acknowledge the challenge in changing attitudes and values, and the emotions involved,
   and spend adequate time talking out proposed changes.

6. Obtain support for the development of the system from high-level administrators
   Lay out a clear rationale to upper-level administrators before finalizing
   recommendations.

7. Ensure that the system is flexible
   Faculty, along with administrators, will have attained a clear and mutual understanding
   of evaluation procedures and policies regarding how the university-wide system will
   apply uniquely to their particular unit.

8. Ensure that the system is legal
   Consult with a university lawyer.

9. Define major faculty responsibilities at the beginning of the evaluation period
   Faculty, along with administrators, will have attained a clear and mutual understanding
   of evaluation procedures and policies regarding weighting of teaching versus other
   responsibilities particular to their unit.

10. Define major faculty subresponsibilities at the beginning of the evaluation period and
     determine their weighting
     Faculty, along with administrators, will have discussed which aspects of teaching (or
     attributes of an effective teacher) will be evaluated in their unit.

11. Define the sources of data to be used to evaluate each subresponsibility at the beginning
     of the evaluation period
     Stipulate a minimum number of respondents for a valid report. Develop procedures for
     reporting in instances of low participation rate or small course enrollment. Caution
     faculty and administrators against interpreting single items versus overall means.
     Describe the role of ratings versus comments in guiding professional development.
     Recommend a schedule of survey administration (number of courses per term).
12. Use multiple sources of data  
   Require multiple sources of data in the faculty evaluation process. Recommend opportunities to correct errors and add relevant information prior to finalization of any personnel decision.

13. Ensure that the data/measures are technically acceptable, i.e., are reliable and valid  
   Implement a research component for the university-wide system to track reliability and internal validity of student input over time.

14. Specifically define the criteria and the standards for each subresponsibility  
   The university community will have discussed the standard for acceptable student ratings (e.g., is "3" acceptable on a 5-point scale?).

15. Train the evaluators to evaluate  
   Recommend that faculty discuss the meaning of the student survey to their students.

16. Train the supervisor in giving feedback  
   Supervisor training will have been implemented on how to communicate with faculty in a constructive manner regarding their annual performance review.

17. Maintain appropriate confidentiality  
   Stipulate that personnel decisions remain confidential. Defer consideration of widening access to student input data until there is a receptive campus-wide climate and adequate university-wide research data indicate a readiness for such discussions.

18. Reward effective performance  
   Avoid the false impression that the system of student input can make finer distinctions than the reward system can support.

19. Combine development with evaluation; have an on-campus consultant  
   The university will have enhanced opportunities for faculty to improve teaching effectiveness by assigning appropriate resources to support professional development.

20. Review the system periodically  
   Stipulate that a 3-5 year cycle of system review occur and that revisions be made accordingly.
IV. RECOMMENDATION FOR A UNIVERSITY-WIDE SYSTEM OF STUDENT INPUT ON TEACHING EFFECTIVENESS

A. Overview and rationale

After review and consideration of (a) the faculty feedback and (b) the research findings and best practice recommendations on student ratings, the Task Force developed a set of principles and guidelines which were then applied to the selection of a survey tool/system and delineation of a process. In summary, our principles and guidelines included the following:

1. Principles for developing a system of student ratings
   - The purpose will be both summative and formative
   - As a **summative** tool, the system will constitute one source of evidence that can contribute to the faculty evaluation process, but is not expected to serve as the sole means for evaluating teaching effectiveness. It will focus on areas in which students can validly judge. It will be inclusive of variables affected by student characteristics. It will guarantee student anonymity. It will preserve the confidential nature of the faculty appraisal process.
   - As a **formative** tool, the system will be both diagnostic and flexible. To be diagnostic, the survey will include specific, behavior-focused items and will be geared toward guiding efforts to improve teaching effectiveness. To be flexible, the survey must be able to accommodate various course goals and varieties of teaching methods.

2. Guidelines for developing a system of student ratings
   - The survey will incorporate the necessary analyses to account for known effects
   - Students will be educated about the purposes and uses of survey results
   - Results will be interpreted in light of multiple sources of evidence of teaching effectiveness
   - Results will be used to foster ongoing professional development and improvement (per E7.0)

3. Selection of a survey system
   The Task Force conducted an Internet search for available commercial systems and followed up via in-depth telephone interviews with representatives of several promising contenders. We deliberated the strengths and weaknesses of each system and narrowed the field to two: IDEA Center and SmartEvals from GAP Technologies. Characteristics of each system are summarized in the table on the following page.
## Table of Comparative Features of Student Surveys Selected for Trial Use

<table>
<thead>
<tr>
<th>Criterion</th>
<th>IDEA Tool</th>
<th>SmartEvals</th>
</tr>
</thead>
<tbody>
<tr>
<td>System supports summative &amp; formative purposes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Built on strong research base</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Developed to facilitate teacher improvement</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Can connect to Wallace Center teaching effectiveness programs</td>
<td>Yes</td>
<td>Yes (not auto)</td>
</tr>
<tr>
<td>Significant faculty time commitment for course set-up</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Significant student time commitment for survey completion</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Reported average response rate</td>
<td>75%-80%</td>
<td>72%</td>
</tr>
<tr>
<td>Large comparison data set</td>
<td>Yes</td>
<td>No (but has item bank w/o stats)</td>
</tr>
<tr>
<td>Customizable survey</td>
<td>No (can add)</td>
<td>Yes</td>
</tr>
<tr>
<td>Written response items available</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>System of reminders to students</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Survey delivery direct to students</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Can integrate with D2L</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Procedures for management of low section enrollments</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Score adjustments</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multiple informative reports</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Filtering of survey results</td>
<td>Yes</td>
<td>Yes (but not standard)</td>
</tr>
<tr>
<td>Statistical work (more than descriptive) conducted at host site</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hosting/data warehousing</td>
<td>Yes</td>
<td>Yes + $7000/yr</td>
</tr>
<tr>
<td>Vendor support to faculty, survey administrators</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transparent process</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ready to support university's needs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Meets RIT technical requirements</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Meets RIT functional requirements</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pilot available at no cost</td>
<td>25 classes</td>
<td>100 sections</td>
</tr>
<tr>
<td>Estimated cost per year</td>
<td>$50,000</td>
<td>1 yr: $22,995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 yr: $19,995 (with hosting)</td>
</tr>
</tbody>
</table>
The strengths of each system include:

**IDEA Online.** This system is from the IDEA Center. Their name, IDEA, is a mirror of the concept behind their system: Individual Development and Educational Assessment. IDEA is a national resource for research on student ratings of instruction. IDEA surveys are based on documented evidence of validity and reliability. Their group summary reports provide aggregated data across courses, which can be used to formulate local norms and/or to identify faculty development needs. There is a national IDEA Tool database built from 3 million forms processed annually from 340 colleges/universities. Data warehousing is included in the cost.

The IDEA system is full-service system, based on a Student Learning Model – the interplay among "divergent methods and objectives, and shifting circumstances." This might be considered the "Cadillac" of all systems in that it plumbs for course objectives and methods, as well as student learner characteristics, and provides clear pointers to professional development recommendations.

IDEA offers two forms. For the Diagnostic Form, faculty identify teaching methods and learning objectives of course. Scores are adjusted for student motivation, work habits, effort, course difficulty, and class size. The form includes 20 items to rate the instructor, 12 items to rate student progress on the learning objectives, 3 items to describe the course, 12 items for students to rate their own characteristics, and 1 open-ended textbox. The Short Form contains only 20 items and can adjust only for student motivation and work habits and for class size. Administration is via any of four electronic delivery options available to students. Response rates are monitored real-time and automatic reminders can be sent to non-responders. The system is focused on providing information to guide improvement via a Diagnostic Report.

**SmartEvals.** This system is from GAP Technologies, the vendor recommended by the 2010 Task Force, which has since implemented some upgrades to their system. The strengths of SmartEvals include the fact that it can be considered a "Do-It-Yourself" System. That makes the system flexible and customizable for each unit on campus. Customers either create their own items, or choose from SmartEvals' item bank built by their 270 users. The latter method would allow for comparison data (but with undocumented statistical properties). The demonstration system appeared to offer an easy path to creating and organizing surveys.

SmartEvals provides various electronic delivery options to students and automatic reminders can be sent to non-responders. Response rates are monitored real-time. The system makes analytic reports available to each user, with "drill-down" capabilities, but the customer (the university) is responsible for filtering results and managing data aggregation (e.g., for small class size). Data warehousing is available for an extra charge.
4. Formulation of a recommendation
As the Task Force considered the relative strengths of our two finalists, we found ourselves unable to arrive at a consensus. Although the Cadillac system (IDEA Tool) was attractive, fully in line with best practice, could our campus community bear the cost? Would faculty balk at completing the Faculty Input Form, however simple or convenient to use? Would student reflections on course objectives and learning outcomes create conflict with the requirements of external accreditation bodies? Will the resources be available to implement a responsive professional development program? Perhaps "Most Flexible" (SmartEvals) will meet our needs better than "Best Practice" (IDEA Tool) and we will manage without information on the important variables known to influence student ratings.

*Given the weight of these unknowns, the Task Force recommends that the campus examine faculty and student experiences with both IDEA Tool and SmartEvals before making a large-scale commitment to either system. A pilot administration of both systems will be conducted and a final recommendation will be made by the Task Force in consideration of an analysis of feedback from students, participating volunteer faculty, and administrators.*

B. Pilot Phase

1. Pilot phase coordinating team
This team will consist of members of the AAC and individuals from the RIT community who have expertise in areas not represented on the AAC. Needs include expertise from RIT's ITS to monitor functionality of each vendor's system with the RIT system. Also needed is an individual from Teaching and Learning (Wallace Center) to facilitate early planning for future professional development needs. The coordinating team would further develop, oversee, and evaluate results of a pilot study in order to arrive at a final recommendation to Academic Senate by the end of winter quarter, 2012-2013.

2. Vendors
IDEA Center IDEATool: Diagnostic Form
GAP Technologies SmartEvals: Survey to include core items revised from the 2010 Task Force report, based on faculty feedback and recommendations from the literature.

3. Pilot design
**IRB approval.** The coordinating team will prepare and submit a request for the use of human subjects to the RIT IRB. The following action steps are planned, pending IRB approval.

**Selection of volunteer participants.** A total of 50 instructors will be recruited from the nine colleges of RIT. The goal is to make the volunteer sample as representative of the University as possible (e.g. each college will be represented proportionately, and diverse members of the faculty will be included, to the extent possible). Individuals will be targeted who (preferably)
teach two sections of one course, or two different courses, with an enrollment of at least 15 students per section in fall quarter 2012.

Volunteers may be full-time faculty members, tenure-track faculty who will be reviewed for tenure in 2012-2013 (with permission of their dean), or lecturers. Volunteers will be cautioned to check that they have ample opportunity in winter and spring quarters to satisfy any department or college requirement to obtain student ratings with their college's standard survey, as the pilot surveys will not necessarily count toward that requirement.

Process. Participants will prepare for each of the two pilot surveys. For the IDEA Tool, they will complete the Faculty Input Form for the Diagnostic survey, indicating the learning objectives and teaching methods for each course in which the pilot survey will be administered. They may also add up to 20 items to the Diagnostic survey. For the SmartEvals survey, a standard set of core items (listed below) will be used, to which participants may add up to 20 items of their own choosing.

Core items recommended for the SmartEvals survey. The first core item is a student self-evaluative item with a Yes or No response. The chosen item asks about class attendance. It was revised in response to faculty feedback to the 2010 core items on commitment to the course and motivation to learn. Survey respondents criticized those areas as being too fuzzy, too hard to measure, and highly dependent on the personality of the student. The goal of item #1 is to indicate the degree to which the student has had enough opportunity to observe the instructor and, therefore, make a meaningful rating. This will be followed by seven instructor-focused items using a 5-point Likert rating scale. While most of these items represent primarily clearer wording, several also provide a different perspective for the student in response to recommendations in the literature and strong faculty feedback on topics such as motivating students, caring about their learning, and asking for students to judge faculty knowledge. Finally, there will be one open-ended item for the student to supply any additional comments.

1. I attended this class regularly. [Yes or No]
   Likert scale for items 2-8:
   (5) Strongly disagree; (4) Disagree; (3) Neutral; (2) Agree; (1) Strongly agree
2. The instructor enhanced my interest in this subject.
3. The instructor presented the course material in an organized manner.
4. The instructor communicated the course material clearly.
5. The instructor established a positive learning environment.
6. The instructor provided helpful feedback about my work in this course.
7. The instructor supported my progress towards achieving the course objectives.
8. Overall this instructor was an effective teacher.
9. One way in which the instructor can improve is through thoughtful student feedback. Please comment on any issues related to the course or the instructor.
Timeline. The pilot will be conducted in Fall quarter 2012. Analysis of feedback data (see Pilot evaluation, below) will occur during Winter quarter 2012-2013, resulting in a final recommendation to Academic Senate. Preparation for the new system will begin in Spring 2013, for implementation in Fall 2013. This timeframe allows for a thorough analysis of both student and faculty feedback. It will also give ample time for RIT to work with the recommended vendor for a full rollout of the system in Fall 2013. Additional tasks during this period will include revising survey items if SmartEvals is chosen, selecting and training a campus coordinator/liaison with the vendor, and educating the campus community regarding the purpose and mechanics of the new system.

A Business Continuity Plan is under development for 2012-2013 to ensure that current University student rating systems (e.g., Scantron) will be adequately supported during the Pilot phase.

4. Pilot evaluation
After student ratings are completed, faculty participants will receive their generated reports from each vendor. They will then be asked to fill out brief surveys about the IDEA Tool and the SmartEvals system regarding the process, survey items, value, preference, outcome (report of results from the vendor), as well as their thoughts on implementing the feedback. An open-ended question will be used for a simple preference and comparison between the two vendors. Examples of possible evaluation survey items include the following (with a response scale for items 1-3 of Strongly Disagree to Strongly Agree):

1. I found the survey easy to set up prior to delivery to students.
2. I understand the reports delivered from the vendor.
3. I plan to seek information on areas of improvement for teaching effectiveness.
4. Which vendor did you find to be better overall and why? (Open-ended)

Students in each class will be told that they are involved in the piloting of a new student rating system and will be asked to fill out a brief survey after completing the ratings of their instructor. Survey items will be constructed to examine the process and evaluation of the survey items. A recommended survey comes from Iraossi (2006) (with a response scale of Strongly Disagree to Strongly Agree):

1. I understood the objective of the survey.
2. I felt comfortable responding to the survey items.
3. The wording of the survey was clear.
4. None of the items were confusing.
5. I was embarrassed responding to one or more of the items.
6. The items appeared biased.
7. The survey was too long.
8. (Open-ended) What issues in rating your instructor were overlooked with this survey, if any?
5. Process for formulating final recommendation
The Task Force will evaluate faculty and student responses to the vendor surveys. Simple statistical analysis will be conducted to examine differences between the two vendors on the quantifiable measures of the specified parameters (e.g. process, survey items, outcome). The open-ended items will also be analyzed to provide depth to the quantitative reports. While faculty will be asked which vendor they prefer, their other responses will be analyzed separately, as well. This will provide support for the final vendor recommendation based on multiple parameters. The Task Force will then compile this information and present a report to Academic Senate, recommending one vendor based on the results of the pilot.

C. Considerations for university-wide launch

The following are some of the tasks that will require the attention of the pilot coordinating team, the AAC, and university offices assigned to support the student rating system as they collaborate in transitioning the university to the newly selected survey.

1. Identifying and training an on-campus coordinator

2. Specifying administration procedures
Procedures are needed to ensure that anonymity of student respondents is maintained. The student response period will need to be defined, both in length and timing relative to the term (e.g., weeks 13-15, inclusive). A reminder schedule must also be scheduled.

3. Establishing mechanisms to ensure adequate response rate
Strategies for encouraging student participation have been summarized in this report (p. 14) and can be found in the literature. Faculty will benefit from help sessions, tip sheets, and electronic reminders to successfully implement these strategies.

4. Delineating analysis and reporting procedures
How well the new system achieves its summative and formative purposes will be a function of how seriously the University strives to arrive at and maintain a common practice in the interpretation of student ratings and their use as part of the faculty evaluation process. Following best practice, the recommendation is to maintain confidentiality of student input data:

- Ratings summary from summative items to be shared with faculty member and supervisor
- Ratings summary from formative items to be provided to faculty member only; to be shared at his/her discretion
- Comments to be provided to faculty member only; to be shared at his/her discretion

Guidelines for analyzing student rating data can be found in the literature. Variables that may require control or post-survey data comparisons (Benton, 2012) include:
• Student motivation
• Reason for taking the course
• Class size and criterion needed to trigger a report
• Workload of the course
• Difficulty of the subject matter of the course
• Level of course (e.g., lower-level vs. upper-level; undergraduate vs. graduate)
• Academic discipline (e.g., humanities/arts vs. social science vs. math/science)
• Number of courses and number of terms over which student ratings are collected per year
V. RECOMMENDATIONS FOR SUPPORTING A SYSTEM OF STUDENT INPUT ON TEACHING EFFECTIVENESS

1. Revise policy E7.0
The language of the RIT Faculty and Staff policy on Annual Review of Faculty, as it refers to student input on teaching effectiveness, is not in agreement with best practice as supported by the research literature. Revisions are recommended that accomplish the following:
Within Section B, Review Process, add that multiple sources of data relating to teaching effectiveness must be consulted. In section B.5, in support of the distinction between the overall evaluative function of the faculty member's supervisor and the information function of student input, replace the phrase "student evaluations" with "student ratings" or "student feedback."
In addition, state explicitly in this section the summative purpose for obtaining student feedback and emphasize that student input is sought as data regarding only one aspect of a faculty member's evaluation, specifically teaching effectiveness. For example, the revised wording might be:
"Standardized student ratings of teaching effectiveness as established by college policy. For non-tenured faculty, student rating surveys shall be conducted in all sections taught during the period of review. For tenured faculty, student ratings shall be obtained in at least one section of each course taught during the period of review. Student ratings shall not be the sole source of data regarding teaching effectiveness. Other possible data include alumni ratings, peer ratings, personal self-assessment statements, syllabi and other course documents, examples of student work, and teaching portfolios."

2. Standardize use of student ratings across the University
Faculty were clear in expressing their perceptions, based on experience, that student input has received either undue weight or complete disregard as a component of their evaluation. Efforts are needed to standardize the interpretation, weighting, and use of student data relating to teaching effectiveness in personnel decisions. It is recommended that faculty supervisors receive periodic refreshers regarding the faculty evaluation process and that regular monitoring be employed to confirm that the principle of seeking multiple sources of evidence of teaching effectiveness continues to be applied. This expectation, communicated from Provost to Deans and from Deans to supervisors, can ensure the valid use of student feedback to their instructors.

3. Develop a system to provide faculty opportunities and incentives to enhance their teaching effectiveness
Given the formative purpose of obtaining student input on teaching effectiveness, it is incumbent on the University to provide mechanisms for faculty to act on that input. Faculty will benefit, first, from professional expertise to translate the feedback into an action plan for professional development, with clear implications for instructional enhancements. Second, faculty will look to the university for an efficient and effective program where they can gain the needed information and skills. Thus, it is recommended that resources be allocated to satisfy
professional development needs in the area of teaching effectiveness, beyond an annual conference, beyond shared blogs or occasional guest speakers.

4. **Develop a campus-wide informational campaign, including students, regarding the role of students in faculty evaluation and emphasizing a new way to talk about student input**

   There is strong evidence in the ratings literature that the value of student feedback increases when its purpose is clearly communicated to students. Students should also be aware that their input is sought, not as "evaluators," but as sources of critically important data, one type of evidence regarding a faculty member's instructional effectiveness. The success of our major effort to launch a university-wide system will be worthwhile to the extent that this message is shared with students through in-class announcements, communications with and from student government, and in campus-wide information campaigns.

5. **Develop a campus culture to promote seeking EARLY TERM student feedback regarding teaching effectiveness and a system to provide faculty opportunities and incentives to effect changes concurrent with instructional delivery**

   Guidelines from the literature about improving teaching effectiveness indicate that end-of-term ratings are limited in their effect and fail to satisfy one's current students who are unable to benefit from future-term instructional enhancements. The practice of early term student feedback has been used to obtain just-in-time information about students' learning experience, allowing changes to be implemented before it is too late to benefit the current-term enrollment. Students who participate in early-term feedback not only appreciate the opportunity to affect their instruction, but also perceive the instructor in a very positive light as one who is concerned about optimizing student learning.

   Faculty will benefit from the development of convenient mechanisms and sample surveys that provide opportunities to their students, along with workshops in how to interpret and apply the findings. Use of this strategy is likely to increase as faculty supervisors provide encouragement, support, and rewards and as a campus culture grows around the importance of continual improvement in teaching effectiveness. Regular training for faculty supervisors regarding the availability of early-term feedback resources and their potential for enhancing teaching effectiveness may be needed.

6. **Develop a research plan to monitor, document, and report on university-wide student input data**

   Numerous researchable questions arise with the institution of a new system university-wide, suggesting the need for resources to establish and carry out a research plan associated with this effort. For example, are there unexpected effects on student ratings of variables associated with course, student, faculty, or survey characteristics? Does return rate remain adequate over time? What strategies are associated with better return rates? Is there a drift in average ratings attributable to implementation of the new system compared to previous systems used in the
various colleges? Do more students realize the value of their input and more faculty find that there is benefit in the feedback? Do more faculty supervisors consult multiple types of evidence in evaluating teaching effectiveness?

It might be expected that student, faculty, and administrators' perceptions of the purposes, uses, and value of student input will change over time as the new university-wide system becomes more familiar and professional development in the area of teaching effectiveness becomes a routine part of faculty activity. A research plan that tracks attitudes, perceptions, and practices will be necessary to document success in achieving intended goals and to guide course shifts for the future. Findings from such a research plan may also lead to a reexamination of the principles established by this Task Force and a collective decision to modify the guidelines delineated in this report.
REFERENCES


Benton, S. (2012). Student ratings of teaching: A summary of research and literature. IDEA Paper No. 50, Center for Faculty Evaluation and Development, Kansas State University, Manhattan, KS.


Cashin, W. (1996, January). Developing an effective faculty evaluation system. IDEA Paper No. 33, Center for Faculty Evaluation and Development, Kansas State University, Manhattan, KS.


The IDEA Center (2004, August). *Using IDEA results for administrative decision-making*. The IDEA Center, Manhattan, KS.

APPENDIX A

Outline of Supplemental Guidelines Appended to the Task Force Charge

The NTID feedback document contained a list of 14 recommendations in developing an online student evaluation/input model. An abbreviated list of those recommendations is:

1. Basis for the Proposed Survey System
2. One Survey for Four Purposes
3. "Evaluation" or "Input"?
4. Focus on Course vs Instructor
5. Formative Feedback as a Guide to Professional Development
6. Mistaken Extensions of Student Input
7. The Likert Scale
8. A 3-Year Data Window
9. Limited-Scale Trial and Initial Assessment Plan
10. Campus Communication Plan
11. Research Plan – A Reflection of Purpose
12. Comparison Groups
13. Small-Group and One-to-One Interactions
14. The Language of Survey Items

The senate taskforce should concentrate on these fourteen items, paying particular attention to items numbers 1-10, and especially number 14. And finally, an issue that should be foregrounded in this entire process is who amongst the RIT community should have access to student input. It is clear that the majority of faculty feedback is very much opposed to sharing this input with all but department heads and deans.
APPENDIX B

Executive Summary of Written responses to the 2010 Task Force Report

I. Clipboard Survey Responses to the Recommended Core Items

1. "I had a strong commitment to this course"
   Commitment not necessarily tied to ability to evaluate—especially in gen-ed requirement
   More useful to understand why the student took the course
   Supplemental questions about attendance, lateness, handing in assignments, etc.

2. "The Instructor motivated me to learn"
   Student has to supply the motivation!
   Some motivation can be negative
   Does not take into account cultural differences among students

3. "The instructor was organized and prepared"
   Students have to get used to all personality types—even disorganized!
   Two questions rolled in to one—separate them, or choose

4. "The instructor communicated clearly"
   Bias towards foreign instructors?
   Revise; many types of communication
   Replace with effectively

5. "The instructor effectively demonstrated knowledge related to this course"
   Why not use “presented” instead of “demonstrated?”

6. "The instructor cared about my learning"
   No real value as a question—student should care about learning!
   Add question about transparent grading metric
   Students should not be asked to judge instructor’s mental state
   Instructors care in a variety of ways, not always evident to the student
   Discriminates against instructors who are emotive or demonstrative
   Perhaps, “shows compassion when I don’t understand”, or similar phrasing

7. "The instructor evaluated my work in a fair and useful manner"
   Some students need extra help, and this can be construed as unfair—eliminate “fair”
   Literature shows students not effective in evaluating this—reflects grade expectations
   Another double-barreled question
   Replace fair and useful with timely evaluation, constructive, etc.
8. "The instructor was very effective"
   More clearly defined: more effective at what?
   Why not drop the “very?” Effective should be the highest rating.
   Students may interpret effective in terms of how well they did in the class!
   Many courses students will not find effective until after college!
   Same rating scales to avoid confusion
   Subjective questions will be used objectively is a problem
   Effective perhaps not understood by ESL students—too vague a concept?
   Change to reflect subtleties—slightly more effective than just the textbook, e.g.
   Only upper class students have had enough instructors to compare effectiveness

9. General comments
   Title instrument “course evaluations”
   Online evaluations ensure that only the positive and negative responses get heard
   Provide revised questions that NTID students can understand (survival English skills)
   Many factors that bias these responses from students: age, sex, required course, etc!
   Questions “normalized” for type of course—required, outside one’s major, etc.
   Students not in a position to evaluate demands placed on them
   Multiple-choice effectiveness is the propaganda of social scientists
   Evaluations are fine, but not their wide dissemination and use
   These questions just measure “likeability” and will lead to grade-inflation and cynicism
   Grade expectation might provide some context for individual evaluations
   Get rid of adverbs in the questions, “very,” e.g.
   Anticipation of lower response rates online
   No questions about facilities, course-specific things, etc.
   Data are meaningless in response rates are too low
   Responses should not be coerced
   Why not use internal resources to do online course evaluations?
   Link evaluation to received grade—problem of letting W students participate
   Questions must apply to everyone—even support faculty for whole 6 is awkward
   Questions do not reward innovation in teaching, but teachers who maintain same syllabus
   Grade distribution for a course would be a useful measure
   Something in the survey to address bias against female instructors?
   Perhaps frame questions as “compared to other classes, (how much did you learn)”
   Definitely have a general comment area
   Numbers based evaluations lead to the path of a vocational-style education system
   Why not leave it to deans, chairs to do qualitative evaluations?
   A more objective set of questions: “were learning objectives clearly stated,” etc.
   Questions aimed at perceptions of faculty, and not at the course
   A question about an instructor being available outside of class?
   Most students not given formal training in providing constructive feedback
No attempt to gather feedback on course material
Must be enforcement mechanism for completion to get a good response rate
Perhaps ask students to report GPA with evaluation
Online comments seem careless or disengaged
9th and 10th weeks the worst times to fill out these things—why not later?

Evaluations will be only one among many sets of data used in evaluating teaching effectiveness
Responses compared only against those from similar courses
Not true in experience!
Administrators should be required to use other sources to evaluate
Administrators will use this as a shortcut rather than observations, etc.

Faculty supervisors should have access to the same data in the survey
Student response rate increased if they know that supervisors see the data
Supervisors should have access to grades of students who fill out the evaluations
Supervisors should collect all of a teaching portfolio to evaluate teaching
Allow access to peer-evaluators
Only information available generally should be made available to supervisors
Chairs should only have access to summative data
Chair should have individual data, but not deans
Online evaluations biases against professor, so shared only with chair
Access to everything will ensure that professors limit optional questions
Made available to promotion committees
Students not responsible for comments, so they will write anything, no matter how untrue
Provide mechanism for faculty to attach comments to raw data
A face-to-face meeting with chair and faculty member should accompany the stats

General Comments
No anonymous comments should be used in evaluations
Supervisors must be trained to interpret evaluations
Distinction between core and optional items: one for promotion and one for development
Add “How well did course prepare you for future employment?”
Anonymous comments can skew evaluations and are usually from extremes
Anonymity breeds lying and slander
Think of ways to get responses form the middle tier
No real way to compare across colleges
Not fair to compare required courses to electives
Single comments often outweigh the rest of the comments
Meaningless unless students fill out evaluations at the same time

**Maximum timeframe during which data would be accessible, and community access**
Humans are nosy—this could cause problems!
This should also apply to provost, deans, etc.
Only accessible by management
Will this lead to popularity cults?
Only if names were attached to evaluations
Professors will not try new things in the classroom
Would be divisive
Students do not have access to other students’ GPA
How would access improve classroom learning?
7 years to correspond with tenure
Should be used only for awards, etc. Too much backbiting in academia for this!
Can’t imagine accessibility in the private sector
Limit accessibility to response rates of 85% or better
No published data until after three years, but keep it permanently after that
Permanence allows the buildup of a reputation
Accessibility will lead to grade inflation
Why not give wider access to other metrics as well, then?
Visible only to direct administrators: dean, chair.
What happens to 50% of faculty members who are below average?
Run assessment for a year’s trial before making decision about accessibility
Available for one year only
Can punish new faculty members for three years for one rough year

**II. Email Exchange Comments to the Task Force Report (only if comments differ from above comments)**

Faculty were polled about questions, but not whether the whole exercise was a good idea
How to account for evaluation of female faculty in the STEM disciplines given the biases
Ask questions such as: are instructors abusive in classrooms; couple them with observations
Are these course evaluations or teacher evaluations?
Include open-ended questions in the instrument
We should be going externally to develop an evaluation system
Do not allow supervisors to use a single quarter’s evaluations against a faculty member
Consult more broadly the relevant literature
Don’t use evaluations for tenure decisions without other corroborating evidence
How to determine is responses are statistically significant?
Why the “paucity” of references? See AAUP discussion
Postpone new evaluations until 2013
More student involvement
Only two questions for instructors: helped me learn, and was she inspirational?
Evaluation parties possible!
Needs carrots and sticks to get good response rate
Students not in a position to know good pedagogy
Even a well-designed online survey will remain compromised in its value if the survey is not strongly supported at the local level of the department; if it is not regularly open to further input and refinement by faculty committees; and if it is not explicitly tied to institutional practices that consistently foster a culture of ongoing faculty development. (Anderson, Brown, & Spaeth, 2006).

Based upon the literature review as well as the findings of this study, online evaluations are most effective when faculty and administrators “buy in” to the process, and when focused efforts are directed at encouraging student participation (Fike, Doyle, & Donnelly, 2010).

Student evaluations of teaching effectiveness are multidimensional, reliable and stable, primarily a function of the instructor who teaches a course rather than the course that is taught, relatively valid against a variety of indicators of effective teaching, relatively unaffected by a variety of potential biases, and are seen to be useful by faculty, students, and administrators (Benton & Cashin, 2012).

Several researchers in the field of teaching and learning have commented that the number of studies of student rating of instruction far exceeds those of virtually any other subject within the discipline (Benton & Cashin, 2012). This is partially reflective of the fact that almost every college or university uses student ratings to collect feedback about teaching effectiveness, but more to the point, it is reflective of the fact that many schools use student rating of instruction as the primary metric for teaching effectiveness when determining tenure, promotion, and merit. That student ratings can have a tremendous impact on one’s career is not something that any university can take lightly when considering changing the way that it approaches gathering the data. This is the more so since there is a great deal of misunderstanding about the effectiveness of student ratings. In a 2005 study, Ronald Berk notes that a comprehensive review of faculty complaints about student ratings yielded almost no statistical evidence to support the majority of complaints and apprehensions (Berk, 2005). Marsh, writing as early as 1984, found student ratings to not only be reliable and valid, but also that students tend to consistently rate the instructor rather than the course (Marsh, 1984). Bearing in mind faculty misgivings on the one hand, and the fact that the literature tends to be fairly consistently supportive of student ratings on the other, this literature review will aim to evaluate the scholarly studies over the past two
decades of student ratings in general, and online student rating of instruction in particular, especially in the context of faculty feedback at RIT.¹

A review of schools that have made the switch to online rating of instruction, as well as a review of the literature about student ratings, reveals that one of the primary reasons for making the switch is for cost effectiveness. Brigham Young University, one of the first schools to adopt an online system of student ratings, found that the cost of student ratings dropped by fifty percent (Univ. of Michigan, www.umich.edu/~eande/tq/faq.htm). Even taking into account the annual fee associated with an outside vendor such as the two that this taskforce is putting forward, namely Idea Center and SmartEvals, online ratings are more cost effective as a whole when the corollary costs of administrative work to type written responses is added to the cost of the physical paper and pencils for over 17,000 students. RIT has a somewhat different, although perhaps related, motivation: the system that RIT uses currently is no longer supported, so that if and when the technology fails, the institute will be left with no alternative but to scramble for an online alternative, or purchase another paper and pencil system, which is certainly available but, as this literature review will demonstrate, not necessary and not cost effective.

There are several advantages to an online system aside from cost, some of which has already become evident as various units on campus have switched to the Wallace Center’s Online Course Evaluation system. Faculty are able to receive their feedback almost immediately after the term ends, which means that for faculty teaching the same class the next term, they are able to integrate student feedback into their courses faster rather than having to wait several weeks, if not months, before receiving the last term’s paper and pencil forms. Another advantage is that research has shown that students are willing to write much more in the open-ended portion of the survey than they are in class (Anderson, Cain, & Bird, 2005). Moreover, at schools in which online ratings have been tried, students are overwhelmingly in favor of doing ratings online rather than in class (Dommeyer, et al, 2004). The one area of concern, which will be discussed below, is that online student ratings tend to elicit lower response rates than in class surveys, although there are a number of non-incentive strategies that could be employed to increase response rates to about the same level as in class surveys, and some schools whose specialty is technical education have achieved response rates higher than those of in class surveys (Layne & DeChristoforo, 1999).

One of the biggest concerns from faculty about online student ratings is that allowing students to complete the surveys on their own time, outside of class, leads to what researchers call a “non-response bias.” In this particular case, the worry is that the good students who have had by and large a good experience of the class will not feel the need to fill out the survey.

¹ Faculty feedback in this context refers to the qualitative and quantitative responses to the clipboard survey created in 2011 entitled “University Wide Student Evaluation of Courses and Instructors recommendations” as well as the rather extensive email exchange that took place over the course of April 2011.
leaving a majority of students who have particular gripes against the professor to vent their rage on the online rating survey. This, the concern goes, will lead to increased negative results quantitatively and a plethora of negative comments from disgruntled students, leading to more negative scores overall. Most of the literature indeed indicates that this is one of the primary worries of faculty, and the faculty feedback that the taskforce was charged with reviewing indicates that this is the case at RIT as well.

The research, however, does not support this position. The vast majority of studies indicate that switching to online surveys did not statistically affect the quantitative results of the survey (Wode & Keiser, 2011). Quite the opposite: the data reveal that the average scores on the survey actually increase slightly in many cases because the students who performed poorly in the class by and large could not be bothered to fill out the survey, while the students who did well in the class were more responsible and took the time to fill out the survey (Liegle, 2004). This last point should go some way to alleviating a major anxiety of online ratings: that underperforming students who often did not show up to class will disproportionately bias the online survey. The data simply does not support this.

Nor does the evidence support the contention that online rating of instruction leads to more negative comments in the open-ended section. Some faculty are concerned that an online system in which students are free to fill out the survey on their own time will lead to such extreme phenomena as “evaluation parties” in which students vie with one another to see who can come up with the most outrageous comments. Other concerns focused on the phenomenon that students that dropped the course might have access to the survey and might be able to write negative comments, having not even fulfilled all requirements of the course. As to the first concern, the evidence does not support the widespread existence of such practices. That is not to say, of course, that anecdotally this does not happen, but the research is clear that there is no real significant difference in valuation between online and paper ratings of instruction. In short, as Donovan and colleagues state, “Instructors using online formats…are likely to find that open-ended comments will not only be quantitatively greater in number and length, but they will contain more qualitative detail than is likely to be found in traditional evaluations” (Donovan, Mader, & Shinsky, 2006). As to the second concern, the literature suggests that students who drop a course, or who never attend the course and thus do poorly, are generally not motivated to fill out the online ratings.

What the research does bear out, however, is that there are some slight biases in online ratings. For example, sophomores are more likely to fill out surveys than seniors, women are slightly more likely to fill them out than males, and students who did well, or expected to do well in the class, were more likely to fill out the survey (Thorpe, 2002; Layne & McGinty, 1999). In fact, in a major study of online ratings, researchers found that the major variable that affected ratings was student motivation, especially as it to the reason for taking the course. Students who
took the course as an elective gave higher ratings than those who took it as a requirement (Cashin, 1995). Similarly, courses within one’s major received higher rankings than courses outside the major. These biases, however, were relatively small, and the research shows that they can be factored in to the evaluations by asking a global question that gets at a student’s motivation, such as prior interest in the course, or whether the course was an elective or a required course (Benton & Cashin, 2012).

One area of controversy is what is termed the “Dr. Fox Effect,” after a 1976 experiment (Williams & Ware, 1976) in which an actor was coached to deliver a lecture that patently made no real sense but, because the actor was overly enthusiastic and used humor and exaggerated body language, the audience rated the lecture very highly. This demonstrates that non-verbal demeanors can influence even a highly educated audience in how highly they rate teaching effectiveness. While there is no doubt that expressiveness tends to influence overall ratings, Cashin states that, “Nevertheless, making the class interesting as well as informative helps students learn content. Expressiveness tends to enhance learning and does not require control” (Cashin, 1995). Further research has indicated, moreover, that the Dr. Fox effect is really only statistically significant in courses in which the student motivation is low (Benton & Cashin, 2012).

One area of concern is that student ratings in classes that are small in number cannot, by themselves, accurately reflect teaching effectiveness. For example, student ratings from classes of ten or less students should not by themselves be used by administrators as evidence of teaching effectiveness, but rather those ratings should be bundled with other classes, or the same course taught several times should be bundled to provide for a larger sample size. Cashin advises in a 1995 study that administrators should “use ratings from a variety of courses, for two or more courses from every term for at least two years, totaling at least five courses. If there are fewer than fifteen raters in any of the classes, data from additional classes are recommended” (Cashin, 1995). Similarly, if response rates are very low, the results of the survey should not be used for personnel decisions unless the course is one in which ratings from previous iterations of the course can be bundled to make a larger pool of raters (Laubsch, 2006). Clearly, there are some issues which have the potential to effect student ratings of instruction, but the literature suggests that with the proper training of administrators, the proper support staff in place at the university, and a survey that is statistically reliable and valid based on many years of research in this field, student ratings can be an effective method of measuring teaching effectiveness.

One caveat that must be made clear to faculty and administrators at RIT is that student ratings of instruction must be only one aspect in the overall evaluation of a faculty member’s performance in the classroom. Most of the literature recommends that other data points also be used in assessing instruction. Cashin, for example, states that “No single source of data—including student rating data—provides sufficient information to make a valid judgment about
overall teaching effectiveness” (Cashin, 1995). Therefore, it should be routine for at least junior faculty members to be observed by tenured peers within their department who are experienced and successful teachers. This allows for yet another measure of teaching effectiveness and provides junior faculty members with another source of formative data with which to improve their teaching (Berk, 2005). It also provides a department with several opportunities to provide support and guidance to faculty for whom teaching is more of a challenge, rather than leaving the issue unaddressed until personnel decisions have to be made either at the third year review or at tenure.

In terms of summative aspects of student ratings, several researchers state that only a few “global” items are enough to be able to make decisions about promotion and tenure (Abrami & Apollonia, 1991; Centra, 1993; Ory, 1994). Edward Nuhfer notes that many surveys collect large amounts of summative data that in the end play little role in personnel decisions. Thus, citing past research, he states that three global questions, such as the following, yield data about general teaching effectiveness that is valid and reliable (Nuhfer, 1996):

1. Overall, how do you rate this instructor’s teaching ability compared to all other college instructors you have now and have had in the past?
2. Overall, how do you rate this course compared to all other college courses you have now and have had in the past?
3. Rate this course as a learning experience.

One note of caution in the use of these types of global items is that, as Nuhfer notes, the statistical correlation with student learning is positive and significant, but that correlation begins to decline when the sample gets smaller. Therefore, these items, while useful in large sample sizes, cannot really be used for individual comparison. Furthermore, as noted above, global summative items provide only one measure of teaching effectiveness and should be combined with other data to obtain as accurate a measure of teaching effectiveness as possible (Seldin, 1993). Overall, however, the vast majority of studies have found that global summative questions are generally reliable and valid and that they do yield a fair measure of teaching effectiveness.

The question of student ratings and formative data gleaned from such ratings leads to the extremely important issue of teaching support for faculty at RIT. If, in fact, a major use of student ratings is to be for formative purposes, then this taskforce believes that the university must take steps to ensure that the appropriate body on campus, whether Teaching and Learning Services or some other group, offers regular workshops for faculty who would like to improve their teaching based on the formative feedback. Cashin, in a 1996 paper, advises that every campus should make available to their faculty instructional consultants or master teachers “who have zero input into personnel decisions” (Cashin, 1996). It might even be desirable for a designated group on campus to convene to begin thinking about ways to assist faculty in improving their teaching based on student feedback. This, combined with the recent emphasis
on peer mentoring, can be a powerful tool that would allow faculty to use student rating of instruction to improve their teaching throughout their career. It would also be to the benefit of students who would be able to see their ratings used in a larger process of improvement in teaching effectiveness.

Another factor that must be taken into account on the summative side of student ratings is who gets to see the data and how those with access to the data are trained. First of all, it is the opinion and advice of this taskforce that the only people allowed access to a faculty member’s student ratings be the faculty member’s direct supervisor, and if the faculty member is up for third year review or promotion, the tenure or promotion committee, and the provost. The students should not have access to the results, even in an aggregated form by department or college. This is based largely off of the considerable faculty feedback that was overwhelmingly against sharing student ratings with students or with other faculty. However, this taskforce can recommend a three year period over which the issue of sharing ratings with others of the RIT community would be investigated. The literature is mixed on the subject. Some researchers advise against sharing data at all (Cashin, 1995, 1996), and some view the sharing of either quantitative, qualitative, or both kinds of data an effective way of demonstrating to students that their voices are heard, as well as an effective way to increase response rates to the online survey (University Leadership Council, 2008).

Having established the recommendation of the taskforce in regards to access to student ratings, it is also the opinion of the taskforce, backed up with plenty of expert advice, that administrators that are in the position of using student ratings to make decisions about merit or promotion be provided with appropriate training so that they would be able to “read” the data in a meaningful and fair way. This would reassure faculty that their student ratings were not being used simply for the “final number,” but that the data as a whole was being interpreted by someone able to understand the data and interpret it accordingly.

Another major point to be addressed is one of the biggest potential pitfalls of online rating of instruction. Most articles that address online ratings note that when the switch is made from paper and pencil surveys, response rates drop, sometimes drastically. This has led several schools to attempt various measures to increase response rates. One of the “tricks” used is to offer incentives, such as a prize, to students who complete the surveys every term (Murphy, 2012). Another method is to offer students who complete surveys extra credit: Dommeyer and colleagues noted that one quarter of one percent of the final grade as extra credit is enough to induce a similar response rate to paper and pencil surveys (Dommeyer, et al., 2004). Other researchers have indicated that simply sending out two reminder emails is enough to dramatically boost response rates (Norris & Conn, 2005). This issue may not be a large factor here at RIT since some research has shown that schools in which the student body is largely
computer literate have equivalent rates to paper and pencil methods (Layne & DeChristoforo, 1999).

It is the opinion of this taskforce, following other research, that the single biggest factor in increasing response rates to online surveys is faculty buy-in. This takes the form of an item in the syllabus highlighting the expectation that students fill out the survey so that student ratings become an expectation of the course, as well as announcements in class (Norris & Conn, 2005). Additionally, faculty should indicate to their class during the term that student ratings are an important aspect of education. These two factors alone have been shown to produce rates that are equivalent if not better than in-class ratings (Wode & Keiser, 2011). As Anderson and colleagues note, “A diminished response rate to course evaluations has less to do with the method of distribution than it does with faculty and student engagement, particularly since engagement reflects the efficacy of evaluation” (Anderson, Brown, & Spaeth, 2006). And finally, one of the biggest concerns on the part of students, especially with online student ratings, is that the survey be anonymous. Instructors must therefore assure their students that their responses will remain anonymous and that their responses are always disassociated with any identifying data (Fike, Doyle, & Connelly, 2010).

In terms of implementing any online system for student rating of instruction, it is the recommendation, following best practices and the advice of researchers (Cashin, 1996), that a pilot program be utilized at the outset of the process. In our case, this will allow for RIT to pilot two possible vendors and, after data is collected from the pilot, to ultimately make a decision as to the product that best suits RIT. Even without having to choose between two vendors, however, researchers caution that a pilot is a good way to gradually introduce online student ratings to campus. Cashin notes, however, that ratings collected in the pilot should not be used for personnel decisions (Cashin, 1996), so we would advise the participation of either associate or full professors at the university, ideally distributed across RIT’s nine colleges.

As noted at the outset of this report, student rating of instruction is one of the most emotive and contentious issues in higher education today. Partly this is because there are many and sundry misunderstandings around the issue, but also because in our profession, tenure and promotion are existential realities that shape our entire futures in the field. This is the more so because more than a few schools tend to use student ratings as the only metric by which to measure teaching effectiveness. Seen in this light, it is understandable that there should be a fair amount of angst surrounding the issue. It is the hope of this taskforce that online student rating of instruction can be an opportunity for RIT rather than a burden or a divisive issue. If the survey is well constructed and based on empirical evidence as collected over the last several decades; if best practices are observed in the use of summative data for merit and promotion purposes; if there is broad faculty buy-in to the system; and if there is institutional support for translating the formative aspects of the surveys into resources for the improvement of teaching,
then we believe that the use of online student ratings will be a positive for the faculty, the students, and the administration at RIT.

Bibliography


Cashin, William (1996). Developing an effective faculty evaluation system. IDEA Paper No. 33. Manhattan: Kansas State University, Center for Faculty Evaluation and Development.


Liegle, Jens & David McDonald (2004). Lessons learned from online vs. paper-based computer information students’ evaluation system. ISECON 21, 1-12.


McCracken, Bridget & Kevin Kelly (2011). Online course evaluations: feasibility study project plan and draft report for San Francisco State University. [http://www.sfsu.edu/~senate/committee/online-eval/](http://www.sfsu.edu/~senate/committee/online-eval/)


Williams, R. & J. Ware, "Validity of student ratings of instruction under different incentive conditions: A further study of the Dr. Fox effect", *Journal of Educational Psychology* 68 (1976): 48–56.


Yale University: Frequently Asked Questions about online course evaluation: [http://classes.yale.edu/help/itg/nce/faq.htm](http://classes.yale.edu/help/itg/nce/faq.htm)