Final Report
Task Force on Student Rating of Instruction (SRI)
Submitted to the Faculty Senate
April 2006

Executive Summary

- The UNCA Faculty Handbook indicates that student rating of instruction (SRI) is used to both help faculty improve their teaching and to make personnel decisions.

- Change is needed in how we do SRI. Faculty are interested in seeing improvements in both the SRI instrument and the overall system of teaching evaluation.

- We designed and tested a revised SRI form which appears to offer improvements over our current form. Additional work is needed to refine the SRI form; this work should be linked with a campus-wide discussion on learning objectives and other campus conversations related to our Mission and Guiding Concepts.

- In addition to changing the SRI form, we recommend additional changes be made. These include a standardization of procedures for conducting SRI as well as a clarification on how the SRI are used in collaboration with other measures of teaching performance in order to make personnel evaluations.

- A continued investigation into SRI and the general process of how we evaluate teaching at UNCA is necessary, and should be led by the Faculty Senate in collaboration with the Center for Teaching and Learning.

History of the Task Force on Student Rating of Instruction (SRI)

The Task Force on Student Rating of Instruction was assembled by the Faculty Welfare and Development Committee from faculty volunteers in spring 2003 as a result of concerns raised by the faculty about our current student rating of instruction instrument and protocol. The work of the Task Force began in earnest during fall semester 2003 following a collection of faculty feedback about SRI.

Continuing Members of the Task Force on SRI

Leah Greden Mathews (Chair)	Associate Professor	Economics
Mark Boyd	Associate Professor	Computer Science
Archer Gravely	Director	Institutional Research
Bill Haas	Professor	Sociology
Bill Sabo	Professor	Political Science
Connie Schrader	Lecturer	Health & Fitness
Megan Wolfe	Associate Professor	Art

The UNCA Faculty Handbook [3.3.3.1.1 Student Evaluation of Teaching Effectiveness] states, “the main purpose of the student evaluation is the improvement of teaching"; teaching evaluation is also used in personnel decisions.

The instrument that we are currently used for student rating of instruction (SRI) has been in use since the 1980s.

Background

We built our examination of the student evaluation of instruction instrument on the following premises.

1. Some degree of systematic student input is essential for effective personnel decisions. Without it, impressionistic and anecdotal evidence can become the primary forces influencing personnel decisions; this is undesirable. In addition, it is essential to consider multiple sources of information rather than relying exclusively on one measure (whether it is qualitative or quantitative). We thus view the process of gathering perceptions from students on faculty instruction via a survey as one element in a portfolio of multiple sources of feedback used to evaluate teaching effectiveness.

2. It is appropriate for students to have an opportunity to provide their feedback on the quality of their instruction. Just as students receive criticism of their work, so to is it appropriate for them to understand and practice critical evaluation of their teachers. Students need to know that their input on teaching is meaningful and that they have an obligation to fulfill their responsibility and take charge of their education.

3. SRI forms serve two potentially contradictory functions: (1) they are sources of critical feedback which is essential for a university teachers’ self-evaluation, development, and improvement and (2) they provide data for chairs/administrators to use when making performance assessments.

   To achieve the first goal—improvement—the rating forms must solicit specific criticism and encourage students to evaluate critically so that faculty can discover where improvement can be made. When student opinion is requested by administrators for personnel decisions, however, faculty naturally prefer the questionnaire elicit positive and favorable responses. Thus, there appears to be an inherent contradiction in the functions that SRI serves.

4. Our current SRI instrument is methodologically flawed and inadequate for achieving the objectives stated in UNCA’s Faculty Handbook which include the improvement of teaching and personnel decisions. There are several reasons for this, as identified in faculty feedback throughout our process:

   (1) There is little statistically significant variation among SRI scores. While this may be attributable to the “Lake Wobegon effect” (UNCA faculty are all above average) or other factors, the result is that for personnel decision-making, there is very little quantitative data gathered from the SRI that can be meaningfully used by chairs and administrators to distinguish among faculty.

(2) When the instrument produces largely highly favorable results there is no incentive for faculty to work to improve their teaching, because improvement is only likely to provide minor incremental increases in scores which are unlikely to be noticed or statistically significant. Perversely, a faculty who has received high scores on the SRI has a *disincentive* to innovate in the classroom, since experimental pedagogies and activities may yield less consistent SRI outcomes.

(3) Some items on the current SRI instrument are not designed to provide information on the quality of instruction, and are thus ineffective indicators of teaching effectiveness. For example, several items ask about the course (as opposed to the teaching of the course); students frequently provide comments about the course *content* as opposed to the teaching of that content.

(4) The wording of the open-ended items on the current SRI instrument yields responses that are not uniformly useful for teaching improvement; indeed, some of the comments faculty receive in this section are irrelevant in that they deal with physical appearance, bodies, and the like. The result is that faculty waste time sifting through useless and sometimes inappropriate comments to get the feedback that may help inform their teaching practice; this is inefficient for both faculty and chairs/administrators alike.

(5) Anecdotal evidence suggests some students do not take their obligation seriously, perhaps due to time constraints, or over sampling (i.e., completing the same form several times each semester) which can lead to respondent fatigue.

In sum, the sense of the faculty and this Task Force is that SRI isn't functioning as it should on a campus that is dedicated to undergraduate teaching. The current form provides little useful data for evaluating teaching performance, except perhaps for distinguishing the weakest teachers among our midst. They are inadequate for discriminating distinguished effort or documenting improvement. The current SRI instrument is largely unusable as a guide for teaching improvement, and in fact may provide disincentives for faculty to innovate their teaching. As a result, SRI fails in the function that our Handbook indicates is its primary purpose, to aid in teaching improvement.

These principles defined our agenda. Our identification of these issues led us to conclude that an improved SRI process would be both useful and desirable for our campus, and that a meaningful first step would be to ensure that our student evaluation instrument is more effective. To that end, we set out to craft a revised instrument that better achieves the objectives stated in the Faculty Handbook.

Revised SRI Instrument and Pilot Test Results

We began by trying to generate methodologically sound student evaluation instruments. We developed fresh questions and response options that placed more emphasis on student learning and less on the teacher□s performance, appeared to be more valid indicators than the ones on our current form, and would produce more variation in the results. We tried to conceptualize the elements of a good classroom experience and develop items which would

capture those ideas. Our goal was to construct measures that would be more accurate indicators of what teachers are expected to accomplish.

The first innovation was a variation on the current instrument (*Experimental Form 1* contained in Appendix 1). It incorporated a series of questions for students to evaluate their progress on selected learning objectives, followed by items asking students to rate the instructor's teaching. The second evaluation form that we tested was a more radical departure (*Experimental Form 2* presented in Appendix 2). This instrument was designed to test whether or not a "thermometer scale" instrument could yield greater variation in responses by asking students to compare instruction in the rated course with others they have had at UNCA. It was tried because some members of the Task Force were skeptical that any form employing ordinal measurement categories would produce significant variation and thus not alter the current pattern of highly skewed positive results.

We felt strongly that we did not wish to recommend adoption of an alternative form to use for SRI before having some data and experience on which to base a recommendation. To that end, we administered a pilot test of the two experimental forms along with the original in a set of 32 classes at the end of spring 2005 semester. These courses represented a mix of disciplines, course level, and included both required and elective courses. Twelve faculty teaching courses in 15 departments/programs participated in the pilot study and over 500 students representing nearly 80% of the students in the pilot courses were sampled. Appendix 3 includes additional information about the pilot test participants and protocol; the complete set of data generated from the pilot test is available at http://www.unca.edu/ir/survey/SRI_Pilot_Rpt.pdf.

While the pilot test did not yield dramatic results, or clearly indicate a "correct" form to be used for SRI, it did produce several insights from the faculty participants:

1. A majority of faculty participants like the idea of asking students to evaluate their progress on learning objectives.

    - Several faculty noted that it would be useful to learn how students perceive their learning in the course as a tool to gauge, in collaboration with other measures, their own teaching effectiveness and thus inform their teaching practice.
    - It was noted that asking students about their learning may be particularly useful in the development of courses that are designated as ILS-intensive, for example.
    - On the other hand, some faculty noted that not all the items were appropriate for their course which led to some student confusion about how best to answer the question.

2. A majority of faculty participants perceived that student responses to *Experimental Form 1* would be more useful to them than responses from our current SRI form or *Experimental Form 2*.

3. A majority of faculty participants recommended adopting *Experimental Form 1* as a replacement to our current form. There was strong opposition to using *Experimental Form 2.*

Conclusions and Recommendations

A. Recommended Changes in Basic SRI Procedures

In order to reduce error and noise in the evaluation process it is necessary to systematize gathering student evaluation data. The steps recommended below are minor and will not correct the basic problems we examined, but they will provide a more formal protocol for the administration of the evaluations and will help minimize the potential for bias that is potentially introduced without a standardized protocol for survey implementation. These recommendations are consistent with generally accepted standards of practice (*see references*). We recommend:

1. Evaluation instruments be administered in the last week of class, to insure that students have the maximum amount of information and experience upon which to base their judgments.

2. The evaluation instrument be given at the beginning of class to help alleviate respondent fatigue and to encourage thoughtful responses.

3. Current policy requiring that full-time faculty members gather ratings from one-half of their courses be reexamined. Several issues were raised as part of our investigation that should be researched further and discussed in a larger faculty venue. These include:
   - The appropriate minimum class size for conducting SRI. A suggestion of a minimum class size of six students was made, but the Task Force reached no consensus on the issue. A related issue is which class size number to use (size after drop/add versus size after the withdrawal date).
   - A rule establishing that the standardized SRI form only can be used in those courses where the bulk of the work is done in a collective classroom setting, and the prohibition of the standardized SRI form from being conducted in independent courses and small research seminars. Courses where a majority of instruction occurs in a one-on-one tutorial situation or in more individualized mentoring scenario should be rated by an alternative protocol best developed by the individual faculty member in collaboration with her/his department chair or the Center for Teaching and Learning.

4. Instructions to students be clarified and explicitly stated *on the form* that students receive, including the steps taken to provide protection from retaliation and anonymity. [These include that SRI results are not made available to the instructor until after grades for the course are submitted, that all open ended comments will be provided only in typed format, etc.]

5. A mechanism be developed for reporting and addressing any suspicion of fraudulent or otherwise inappropriate activity that may arise during the SRI process, including clear instructions for how staff should deal with these situations so they are not left to make decisions that may impact the SRI

process. Situations reported by faculty and staff during our study that suggest the need for this type of policy include:

- Individual/stray forms submitted after the course packet has been received in the Office of Institutional Research
- A larger number of completed forms included in the class packet than there are students enrolled in the course
- An excessive number of erasures on the forms, and
- Indications that faculty members have prompted students to respond in particular ways in order to achieve a desired outcome.

B. Barriers to Designing and Implementing a Superior SRI Instrument

Our discussions, faculty input, and the pilot test helped us clarify several critical issues that need to be addressed if UNCA's SRI practices are to improve.

1. Logistical obstacles must be overcome. It appears that using the same SRI form for all disciplines is unwise, except for perhaps a small number of global items. Teaching styles and learning objectives are as diverse as the disciplines taught and a unified form may necessarily contain items that have no meaning to students in some courses. This creates confusion and contributes to less than careful responses. The university must be willing to commit resources to implementing an evaluation process with different forms tailored to different disciplines if substantial improvement in SRI outcomes is desired.

2. The current sampling procedure biases results, and must be addressed. The small and unrepresentative nature of respondents in some courses make statistical summaries derived from the current form effectively meaningless. This is particularly a problem in upper division classes with small enrollments. Reliability assumes that all students respond to the questions in the same way. Clearly this is not the case, but the assumption is reasonable if the deviations are randomly distributed across the student body and thus cancel each other out. With a large population, this assumption may be warranted but it becomes impossible to defend when students are distributed selectively in courses.

3. It is possible (some of us suspect likely) that students are not fulfilling their responsibility of being critical evaluators which might account for the lack of variation in responses. We did not investigate this issue, but ad hoc evidence suggests that some students complete the SRI forms with little thought. There are several possible reasons including student fatigue at filling out the same form again and again, a belief that their responses do not matter, the limited number of response options available to them, and a desire to minimize the amount of time devoted to the task.

C. Continued Investigation

We urge the Faculty Senate to continue to investigate ways to improve the SRI instrument and process. Our thoughts include the following:

1. While neither of the alternative instruments we developed and tested is ideal, a majority believe that *Experimental Form 1* moves us in the right direction. In order to continue the investigation into improving the SRI instrument, a majority supports adopting a revised version of *Experimental Form 1* for a trial campus-wide adoption.

   Two members of the Task Force are not enthusiastic about this recommendation because they are not convinced that Experimental Form 1 produces noticeably different or more valid results than the current instrument. They are willing to endorse a trial of the alternative form, however, because they are dissatisfied with the status quo and are interested in continuing investigation into improving the SRI process. We imagine that this sentiment may be echoed by other faculty, based on feedback we have received about our current SRI form.

2. One advantage of *Experimental Form 1* is its relationship to UNCA's Guiding Concepts. Its emphasis on student learning offers a potentially useful foundation for future discussions about teaching and learning at UNCA. Moving toward a student-centered evaluation instrument that attempts to measure how our classroom teaching helps students *learn* appears to be consistent with this approach.

   It might be meaningful, for example, to merge the process of student rating of instruction with a campus-wide discussion of learning objectives. This would ensure that the link between our teaching (input) is linked with student learning (output) in a more meaningful way, which should facilitate other campus objectives.

   Continued investigation might profitably begin with gathering input from students about what they consider when evaluating a teacher and formal inquiry into how students respond to the questions. This information could be used to increase the reliability of the instrument and inform procedures for how to select courses or students to evaluate teaching. This type of information is mandatory if the evaluation process is to place greater emphasis on student learning.

3. The open-ended questions on the reverse side of our current SRI instrument should be reworded but not eliminated. We did not pilot test alternative wording of these questions in order to avoid student fatigue; it is clear, though, that rewording them so they are more clearly designed to give feedback that will be useful for faculty in improving their teaching practice will be beneficial. Suggestions that we gathered from faculty for alternative wording appear in Appendix 4.

4. Since we suspect that several of the procedural and resource problems could be resolved if the university adopted an online rating system, it would be useful to experiment with an electronic rating process. Having students complete SRI online would reduce the burden on program assistants who are currently required to type student comments and translate often illegible handwriting. A pilot test would be helpful in identifying the problems with such a process and whether

issues such as anonymity and response rates could be adequately addressed. While there would be upfront costs associated with a transition to online SRI, it is likely that there would be several benefits:

- Program assistants will spend less time typing the responses to open-ended questions. At the end of each semester, hundreds of hours of valuable labor is spent typing the comments, which are often difficult to read.
- Faculty could gain access to feedback more promptly (via automated delivery after grade submission), which may facilitate suggestions for improving their teaching.
- Having SRI responses in an electronic form may facilitate the work of chairs/program directors and the work of the promotion/tenure and teaching award committees via the simplification of the collection of multiple semesters' worth of data. In addition, electronic viewing of SRI by these groups may facilitate the confidentiality of the results, which currently can be inadvertently mishandled due to their transmission via hard copy.

D. SRI and Personnel Decisions

We believe continued investigation into a better SRI instrument and general protocol for evaluating teaching at UNCA is essential. For a quality undergraduate liberal arts teaching institution there really is no theoretical contradiction between the functions student evaluation instruments can serve. UNCA has a right, even an obligation, to demand rigorous evaluation of instructors for both professional development and personnel evaluation purposes. Indeed, a secure teaching faculty committed to improving their professional work will expect the evaluation process to do both.

We do not think, however, that UNCA's faculty is sufficiently confident to have student rating serve both a formative and summative function. There are two related sources of this insecurity. The first is the inherent deficiency in the SRI instrument, whether it be the current instrument or another form. It is clear from the literature that there is no one, single, perfect form that can perform both formative and summative functions, largely because it is not humanly possible to expect a single set of student evaluators to be able to provide enough information for instructors to get what they need to improve their teaching while at the same time providing personnel decision makers the kind of information that they must have in order to make informed personnel decisions. Thus, it is unrealistic to expect any SRI instrument to serve as the sole indicator for these two functions.

A more fundamental concern is a lack of faculty confidence in evaluators' ability to make sound judgments on the skewed data the instrument generates. Concern about how the results from SRI forms would be used by administrators were noted by several faculty participants in our experimental study, for example. In addition, a majority of the faculty feedback received throughout our process had a refrain that echoed the sentiment that *it was how the data gathered from the SRI forms was being <u>used</u> that was a more pressing concern than just the poorly worded questions on our current form*. Generally

speaking, the sense of the faculty is that an over reliance on the quantitative data provided by SRI (however effective or ineffective) is undesirable.

Clearly, judgments about a faculty member's teaching must rely on multiple indicators. However, we acknowledge that SRI data provide the most objective longitudinal information available. Finding a superior instrument is necessary in order that this component of personnel decisions is based on appropriate statistical data generated by valid instruments with precise controls on how surveys are conducted.

We urge the Faculty Senate to adopt guidelines governing the use of student rating of instruction in personnel decisions. Currently, it is not clear to many faculty what mechanisms are used to determine or otherwise evaluate their teaching performance. What is the relative importance, for example, of student evaluations in the context of faculty statements, innovative activities, student letters, peer evaluations, exit interviews, senior questionnaires and other means of gathering information? Does this vary across departments or divisions? The current process of evaluating teaching appears to many faculty to be somewhat arbitrary, given the poor quality of our current SRI form and the perceived over reliance on it. A good SRI instrument is a necessary, visible first step to making the evaluation process more consistent, but it is not the only change that should be pursued.

In order to clarify how data from the evaluation instruments will be used, it is essential for the University to develop clear and systematic guidelines as to how the results of the questionnaires will be used in personnel decisions. In the short term it seems reasonable for each department to adopt such a policy for their own use. These could then be submitted to the administration for the development of such a policy which will then be approved by the Faculty Senate.

Appendix 1 Experimental Form 1

Student Rating of Instruction

Course:_____ **Section:**_____

Instructor:_____

Year:_____ **Semester:**_____

Use the following scale to indicate the amount of progress you made on each of the objectives listed below.

1 = No apparent progress
2 = Slight progress; I made small gains on this objective
3 = Moderate progress; I made some gains on this objective
4 = Substantial progress; I made large gains on this objective
5 = Exceptional progress; I made outstanding gains on this objective
6 = This objective is not applicable to this class.

A. Student Progress:

1. Gaining knowledge	1	2	3	4	5	N/A
2. Understanding ideas	1	2	3	4	5	N/A
3. Developing creative capacities	1	2	3	4	5	N/A
4. Developing writing skills	1	2	3	4	5	N/A
5. Developing speaking/discussion skills	1	2	3	4	5	N/A
6. Developing quantitative skills	1	2	3	4	5	N/A
7. Learning to analyze and evaluate ideas, texts, and arguments	1	2	3	4	5	N/A

B. How often did your Instructor do the following?

	Hardly Ever	Occasionally	Some-times	Frequently	Almost Always
1. Explained course material clearly.	1	2	3	4	5
2. Created a classroom environment that encourages students to be actively involved	1	2	3	4	5
3. Displayed an interest in student learning	1	2	3	4	5
4. Was accessible outside of class	1	2	3	4	5
5. Gave assignments that contributed to my understanding of the material	1	2	3	4	5
6. Provided useful feedback on my performance	1	2	3	4	5
7. Helped me develop an appreciation for the subject	1	2	3	4	5
8. Made me think	1	2	3	4	5

C. Overall Assessment:

	Poor	Fair	Satisfactory	Good	Excellent
1. Rating of the instructor	1	2	3	4	5
2. Rating of the course	1	2	3	4	5

What grade do you expect to earn in this course? A B C D F S U I W T

Appendix 2 Experimental Form 2

Course: _____ **Instructor:** _____ **Semester:** _____

1. How many credit hours have you earned at UNCA? _____

2. What is your cumulative GPA? _____

For each of the following questions, enter a number between 1 and 99 using the scale at the right as a guide.

3. Compared to other **CLASSES** you have taken at UNCA...

   (a)...how demanding was this class? _____

   (b)...how much did you learn in this class? _____

   (c)...how much of the required work did you complete for this class? _____

   (d)...how challenging were the class sessions? _____

   (e)...how provocative was this class? _____

   (f)...how much did the assignments contribute to your understanding of the material? _____

   (g) how good was this course overall? _____

4. Compared to other **INSTRUCTORS** you have taken at UNCA...

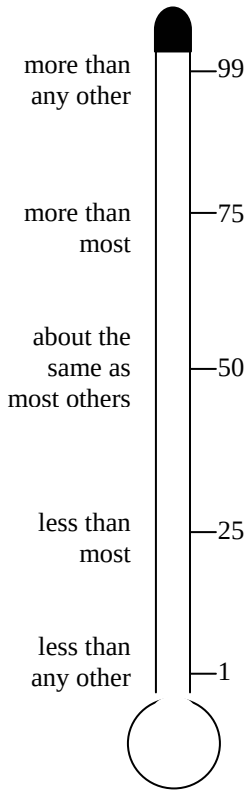
   (h)...how clearly did the instructor explain the course material? _____

   (i)...how useful was the instructor's feedback on the assignments? _____

   (j)...how interested was the instructor in student learning? _____

   (k)...did this instructor make you think? _____

   (l)...how good was this instructor overall? _____



Appendix 3
Outline of the Participants and Protocol Used in the Pilot Test of the Revised Student Rating of Instruction Form

Participants

- 13 faculty volunteers
- 32 course sections
  - Only classes > 10 students
  - No seminar courses
- All divisions represented
- Number of students by division, course level was approximately equal to the population (see pilot test report at http://www.unca.edu/ir/survey/SRI_Pilot_Rpt.pdf for specific comparisons).

Protocol

- Random identifier number assigned to a set of forms to allow us to see if individual students responded to forms differently

- Each student completed a set of forms which included (1) existing SRI form, (2) Experimental Form 1 and (3) Experimental Form 2

- Forms were clipped together and we asked that students completed the forms in the order in which they were presented in order to minimize conditioning their responses

- We asked faculty to conduct the pilot test during the last week of class and at the start of the class session

A Complete Report of the Results of the Pilot Test conducted at the end of Spring 2005 semester is available online at http://www.unca.edu/ir/survey/SRI_Pilot_Rpt.pdf

Appendix 4
Suggested Alternative Wording for Open-Ended Questions

Our recommendation is that a set of open ended questions be included with Experimental Form 1. Several faculty submitted ideas for these questions; these are indicated here for future consideration.

- What did the instructor do that helped you learn?
- What else could have been done to help you learn?
- Which of the *x* specific goals of the course have you made progress in and in what ways? What did the instructor do to facilitate this progress?
- What are your suggestions for improving the course?
- What class activities have been the most useful to help you learn the material of the course?
- What would you like to say about this course?
- Evaluate your instructor.

References

Braskamp, L.A. and Ory, J.C. 1994. *Assessing Faculty Work: Enhancing Individual and Institutional Performance.* San Francisco: Jossey-Bass.

Cashin, W. E. 1995. "Student ratings of teaching: The research revisited." Idea Paper #32. Center for Faculty Evaluation and Development, Kansas State University.

Center for Teaching and Learning, The University of North Carolina at Chapel Hill, "Student Evaluation of Teaching," *For Your Consideration…Suggestions and Reflections on Teaching and Learning*, CTL Number 16, September 1994.

Centra, J. A. 1993. *Reflective faculty evaluation*. San Francisco: Jossey-Bass.

Lewis, K.G. 2001. Techniques and Strategies for Interpreting Student Evaluations. *New Directions for Teaching and Learning*, 87. San Francisco: Jossey-Bass.

McKinney, Kathleen, "What do Student Ratings Mean?" *The National Teaching and Learning Forum*, vol 7, no 1, 1997.

National Research Council. 2003. *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics*. Washington, D.C.: National Academies Press.

The University of Washington, *Recommendations for Use of Student Ratings of Instructors in Merit and Promotion Decisions*. Available at http://depts.washington.edu/cidrweb/FCIQGuide.pdf.

Theall, M. and J. Franklin. 1990. *Student Ratings of Instruction: Issues for Improving Practice*. San Francisco: Jossey-Bass.

Vice Provost for Academic Affairs, The Pennsylvania State University. *Guidelines for Use and Administration of the Student Rating of Teaching Effectiveness (SRTE) Forms**. Available at http://www.psu.edu/dept/vprov/SRTE/guidelines.htm. *These guidelines were developed in consultation with a subcommittee of the Senate Committee on Faculty Affairs and issued in 1987. They were revised in December 1997, and revised again, effective July 1, 2004.