
One University's Measured Approach to Student Evaluations

Dale Carpenter

Western Carolina University

Abstract

Student evaluations of teaching occur at most universities and colleges in the United States and are used for a variety of purposes including course improvement and as data to evaluate instructors. Increasingly, universities manage the collection of student perception data about courses and teaching with commercially available software. This is a report of the work of one university to review the process of collecting and using student assessment of instruction data and to determine how the data would be used. The work of a task force to examine ten years of collected data, review literature, and seek input from stakeholders is presented in a case study format with task force recommendations and a report on subsequent implementation

* Corresponding author: Dale Carpenter, School of Teaching and Learning, Western Carolina University, 106 Reid, 1 University Drive, Cullowhee, NC 28723. carpenter@wcu.edu

Recommended citation: Carpenter, C.D. (2021) *One University's Measured Approach to Student Evaluations*, 2(1), 20-34. doi:10.32473/jahe.v2i1.122007

©Journal of Assessment in Higher Education. Published under Creative Commons License CC BY- 4.0 International.

Introduction

Like many universities, Western Carolina University (WCU), a mid-sized regional institution, has been using student evaluations for many years and transitioned to an online version of collecting student perceptions of courses and teaching in 2008 using a major vendor. For ten years, almost no changes occurred in content and how evaluations were administered. However, the use of student course evaluations became increasingly institutionalized over that timespan particularly in annual faculty evaluations and decisions about reappointments, tenure and promotions. Perhaps the primary reasons to initiate a task force to review the Student Assessment Instrument (SAI) process at Western Carolina University were the concerns expressed by some faculty members that student evaluations were biased, had too much influence in career decisions for faculty, and were not collected using best practices. At the beginning of the 2018 – 2019 academic year, the chair of the Faculty Senate, with the support of the Provost, initiated a 13-member Task Force to examine the process and use of student course evaluations. This is a description of the work of the Task Force including the recommended actions and the response of the university community.

The Task

The Faculty Senate charged the SAI Task Force to:

1. Collect a wide range of research-based evidence and faculty input about best practices in SAI processes at WCU and at other institutions;
2. Using this research and faculty input, review guidelines for making information from SAIs appropriately available to faculty and administrators;
3. Evaluate relevant models and potential vendors; and,
4. Develop a timeline and process for implementation and subsequent ongoing reviews.

Recommendations and a report were due toward the end of April for consideration by the Faculty Senate with interim progress reports during the year. The Task Force was composed of eight faculty members and five administrators/staff members across campus with four females and nine males. Three of the eight faculty members were junior faculty at the assistant professor level. Of the senior faculty, a department head, former department head, former dean, and chair of the Faculty Senate were on the Task Force. Two administrators represented the Office of the Provost and two were from the faculty support center on campus.

Description of the Student Assessment Instrument Used

The Student Assessment Instrument used by WCU since 2008 is called “CoursEval” owned by *Campus Labs*® and is a web-administered instrument with 20 statements organized in five categories that students are to rate on a four-point scale: Strongly Disagree, Disagree, Agree, Strongly Agree, and a Not Applicable choice. Although different forms were developed for different kinds of courses such as labs, online courses, etc., the form most often used is the one for lectures. There is no option for

adding questions. Of the 20 items, 17 begin with the phrase, "My instructor..." e.g., My instructor evaluates my performance fairly.

There are two open-ended questions: 1. Describe the best aspects of this course. 2. Describe changes that could be made to improve the course. Students have a text box in which to respond to these questions. Instructors have had the option to use responses to open-ended questions in their materials and all results are sent to the instructor's department head.

Task Force Methods

Over the 2018 – 2019 year, the Task Force engaged in the following activities:

1. Collected stakeholder input from students, faculty, and administrators using forums, group meetings, informal conversations, and surveys;
2. Reviewed the last ten years of our institutional SAI data to include response rates and possible presence of bias in ratings;
3. Reviewed recent literature, both scholarly and journalistic (e.g., Inside Higher Ed), on student evaluations of instruction; and,
4. Surveyed student assessment software used at other University of North Carolina system institutions and participated in invited web-based demonstrations from four major SAI vendors.

Stakeholder Input

Faculty Perceptions

It was decided early in the process that since the Task Force was created by the Faculty Senate with progress reports due at intervals, the perceptions of faculty would be a major factor. The chair of the Task Force was a faculty member, and the Chair of the Faculty Senate was a member of the Task Force. Therefore, there were three publicized university-wide Faculty Senate sponsored conversations about student assessment. Each conversation was attended by approximately 12 to 25 faculty members and members of the Task Force. The role of the Task Force members in these settings was to facilitate conversation and to listen. In addition, a campus-wide survey was sent to all 520 fulltime faculty members with seven questions about the SAI. Responses were received from 266.

Student Perceptions

Perceptions of students about SAIs were collected by a focused conversation with members of the Student Government Association (SGA), by informal and undocumented conversations with classes by instructors, and by a survey electronically sent to all undergraduates (approximately 10,000) about several topics including three questions about the course evaluations. Directions were – "On a scale of 1 to 10, please rate your agreement with the following statement where 1 is Strongly Disagree and 10 is Strongly Agree."

Administrator Perceptions

A few members of the Task Force met with department heads twice and with the Provost Council once. The Provost Council is comprised of academic deans and administrators in the Office of the Provost.

Review of SAI Data

The WCU Office of Institutional Planning and Effectiveness analyzed the institution's SAI data and found that overall response rates in the first year of implementation, 2008, were 32% and improved to 67% by 2012. The most recent five-year overall response rates from 2013 to 2018 were between 69% and 75% overall. WCU has been offering the incentive of viewing course grades as soon as they are submitted for those completing the SAI on a course since 2008. Although, overall response rates increased generally, reasons for growth are speculative but include familiarity with the process, better email messaging, and instructors encouraging student completion in some areas. Response rates vary by course level, delivery format, etc. The Task Force did not attempt to investigate reasons for uneven response rates.

The Task Force requested an investigation by our Office of Institutional Planning and Effectiveness of WCU SAI data to determine bias regarding instructor overall ratings for gender, age, and race. The last three years of data were examined. The major finding was that only one significant difference at the .05 level in ratings was discovered regarding possible instructor bias - *Female instructors under 40 years of age had slightly higher or better overall ratings than males 55 years of age or older*. This difference does not in itself prove bias, and no other differences were detected in overall SAI ratings for gender, age, and race.

Review of SAI Literature

The available literature concerning student assessment of instruction is substantial, and a few Task Force members individually compiled sources from scholarly literature; news items, e.g., InsideHigherEd, Chronicle of Higher Education, and online resources, e.g., blogs, YouTube videos. While Task Force members collectively and individually reviewed many different resources identified in the References list and shared with each other, most of the impactful findings are included in two major reviews of the literature presented in different formats.

Review of SAI Vendors

The Task Force conducted an informal survey of University of North Carolina constituent institutions to determine the SAI software being used. Based on that survey, four products: Blue, CoursEval, EvaluationKit, and IOTA360, were each demonstrated at different times and dates across the academic year. The purposes of the review of

vendors was to survey how our vendor compared with other popular vendors and to see if WCU was using “best practice” in our SAI process. The questions focused on functionality in the context of administration to students, flexibility to tailor to customers, and reporting capabilities. The Task Force did not investigate back-end technology issues and pricing.

Major Findings

The Task Force learned the topic of student ratings of courses is controversial; beliefs about the value of student assessment instruments seem to vary by stakeholder role; and, few have a comprehensive grasp of the evidence-based literature on student perception instruments.

Stakeholder SAI Perceptions

Faculty

Responses to a survey of faculty members were received from 266 faculty members out of a total of about 520 fulltime faculty members. Finally, as might be expected, Task Force members also reported having many informal conversations throughout the year with colleagues on the subject of the SAI and the Faculty Senate members discussed preliminary findings and provided input a couple of times throughout the year when the Task Force chair provided preliminary progress reports.

During the conversations with faculty, many issues were expressed and the Task Force summarized them in bullet form in their final report as seen below.

- Concern that SAI data are improperly used;
 - Ordinal data are averaged;
 - Overall means are used when items are grouped by categories;
 - Median ratings may be better than means/averages;
- Concern that SAI data carry too much weight in AFE (Annual Faculty Evaluation) and Collegial Review (i.e., Reappointment, Tenure, and Promotion process);
- Concern that students are not asked about their own effort and success;
- Concern that questions do not focus on learning outcomes; most questions are about the instructor;
- Concern that questions cannot be added by instructor or programs;
- Concern that negative and toxic responses to open-ended questions are too easily made and cannot be validated; instructors state that some negative statements are false and present false impression of course and instructor;
- Concern that WCU is not using best practice and expressed eagerness for information about best practice regarding SAIs;
- Concern that other colleges and universities are de-emphasizing SAI data in AFEs and collegial review decisions;

- Concern that SAI data may be biased based on instructor demographics, e.g., age, gender, race, native language;
- Belief that student feedback has a valid purpose; and,
- Affirmation of value in multiple sources of teaching effectiveness, e.g., peer review of materials, peer observations.

Faculty survey results (n=266) revealed that respondents overwhelmingly (90%) believed that students should have the opportunity to provide feedback on course quality and that department heads/directors should have access to SAI data (89%). Over half of the respondents (55%) agreed that SAI data should be considered in Tenure, Promotion, and Reappointment (TPR) decisions. Less than half (42%) favored some form of regular/periodic peer evaluation instruction, instead of the current SAI process. Respondents also acknowledged that they read their SAI results every semester (83%) and used SAI data to improve instruction (59%). About half of the faculty survey respondents (49%) indicated that they also used a self-developed feedback survey in their courses.

Students

In a focused conversation with members of the SGA with a former chair of the Faculty Senate who emailed a summary to the Task Force, students stated that the current process did not fit all courses, and science labs were an example. Students preferred department level faculty evaluations tailored to the discipline and course. Some students wanted midterm evaluations so that a bad course could be fixed before the end of the course. Students said that the evaluations should address learning as opposed to a professor's personality. Some expressed a desire for more open-ended questions and wondered if and how their feedback was used. Reports from Task Force members indicated that the issues raised by the SGA were very similar to the issues discussed by students in other informal conversations across campus in classes, student organizations, other student groups, and individual conversations.

Average ratings from respondents to the student survey (N = 1139) were positive for the following questions on a scale from 1-10:

- I look forward to completing my Course Evaluations (avg = 7.7);
- The process of completing course evaluations is easy to navigate (avg = 8.4); and,
- Course Evaluations offer me an opportunity to provide feedback to faculty (avg = 8.5).

Administrators

Much of the conversation with administrators concerned whether student responses to open-ended questions should be available to department heads and other administrators. Provost council members and department heads acknowledged that student comments could be unfair, inaccurate, and could bias administrators. However,

administrators stated that good administrators were aware of the problems and avoided letting unconfirmed statements negatively affect perceptions and evaluations of instructors. Administrators, both department heads and deans, stated that responses to open-ended questions were important because students expected administrators to read them, and information presented was not likely to be available and open to further investigation unless department heads had access to the comments.

Software Demonstrations

Discussions among Task Force members about observations and perceptions regarding the software demonstrations were summarized in the Task Force report:

- The demonstrated software products appear to be similar from the viewpoint of faculty.
- TF members were surprised with the variety of functions available in CoursEval and were unaware of some currently available features that WCU was not using.
- TF member discussions revealed that many concerns about our current implementation of CoursEval were internal WCU decisions made in 2008 and were not related to the functionality and features of CoursEval, e.g, selection of questions, flexibility to add questions by instructor or program, timing of survey.

Review of Literature

The two most impactful sources of literature on student evaluations for the Task Force were reviews of the literature. Linse (2017) authored an article in *Studies in Educational Evaluation* titled *Interpreting and Using Student Ratings Data: Guidance for Administrator and Faculty Serving on Evaluation Committees* that examines and summarizes the huge body of literature on student evaluations and provides guidelines for using student ratings. In addition, Task Force members were informed by a screencast by Barre (2015) summarizing the literature on student assessments with an update three years later (Barre, 2018).

The reviews found that the topic of student ratings, as well as the terms used to describe student ratings remain controversial. Some terms include student ratings, student evaluation of teaching (SET), student assessment of instruction (SAI), course evaluation, and student perception of instruction. Student ratings of courses and teaching have been around for decades, and the major stated purposes are to evaluate faculty to improve the learning and teaching environment, and to demonstrate accountability of the institution. While there are reports in *InsideHigherEd* and *The Chronicle of Higher Education* that some colleges and universities have eliminated or plan to eliminate student ratings, most continue to use student ratings to some extent (Linse, 2017).

There are claims in the literature that the following factors have been reported to impact student ratings:

- Instructor characteristics such as gender, race, age, native language, presence of a disability;
- Perception of other instructor characteristics such as warmth, approachability, attractiveness;
- Size of class;
- Response rate;
- Grades in class or perception of grades by students;
- Perceived difficulty of class; and,
- Class workload and subject of class, e.g., math, humanities.

The Task Force concluded that the literature revealed that these factors generally exert minimal to negligible impact on student ratings although, as Linse (2017) points out, instructors may be able to point to specific examples of bias.

The literature showed that although the relationship of student ratings to student learning has not been found to be strong and the reliability and validity of student assessments of teaching has been questioned, student ratings were found to be studied much more and were more valid than other kinds of assessment of teaching including peer-observation, focus groups, and external review of materials (Barre, 2015; Linse, 2017).

Discussion

Perhaps the primary reasons to initiate a task force to review the SAI process at Western Carolina University were the concerns expressed by some faculty members that student evaluations were biased, had too much influence in career decisions for faculty, and were not collected using best practices. When the Task Force discovered that the current practice had not been examined since its inception in 2008, Task Force members embraced the challenge and quickly learned several things. Among them, the Task Force learned the topic of student ratings of courses is controversial; beliefs about the value of student assessment instruments seem to vary by stakeholder role; and, few have a comprehensive grasp of the evidence-based literature on student perception instruments.

It was clear from the very beginning of the Task Force work, and did not change during the year, that all stakeholder groups agreed that students should have a means to provide input about courses. The concerning issues for most stakeholders were around our Student Assessment Instrument questions and process and how the information was used. Regarding the instrument and process, faculty members were surprised that neither the instrument nor the process had been changed or reviewed in ten years since the inception of the online administered survey. Also, few knew the history of why and how the current SAI had been developed.

Conversations with veteran faculty members who had been at the university when the current system was developed, including a few who were involved in the SAI development at the time, and a review of the documentation establishing the process

and SAI explained two important facts for the Task Force members. The first was that the University of North Carolina system, of which Western Carolina University is a constituent institution, mandated in 2008 that each university must have a standard student assessment system for teaching. In addition, the purpose was widely interpreted at that time to mean that the focus of the instrument was to assess instructor performance rather than the quality of the course experience or how students perceived what they learned in each course. Consequently, the overwhelming number of questions in the WCU SAI implemented in 2008 began with the phrase “My instructor...”

Faculty and student complaints that the instrument failed to focus on what students learned and the quality of the experience were better understood when it was known that the original purpose did not include those things. Students and faculty members agreed in our Task Force work that the instrument seemed to be overly and inappropriately concerned with what some termed “instructor personality.” Therefore, Task Force members agreed that questions about the course experience from the student perspective should include items that focused on students’ perceptions of their effort, what was learned, and course implementation to include organization, clarity of expectations, and the climate of the class regarding inclusivity or respect for all students. The Task Force recommended possible questions for a revised SAI.

Some individuals among the faculty suspected that concerns about the content of the questions and the rigidity of the process as evidenced in not allowing changes in questions or adding questions based on program or delivery format were due to the software selected in 2008. The Task Force learned that the software being used, i.e., CoursEval, and other popular commercial software reviewed allowed for virtually all functions that were suggested as desirable including the ability to change or add questions by program or course and administer the SAI more than once and at different times. WCU had simply not used those functions but could if desired; therefore, the Task Force made no recommendations regarding software.

When the Task Force collectively considered the allegations that SAI data were biased, they found literature that appeared to confirm bias. However, Barre (2015, 2018) and Linse (2017) review and address these allegations and state that most of the studies have serious methodological issues. They both contend that bias effects are not consistently found in sound studies and have minimal impact on student ratings, including:

- Instructor characteristics such as gender, race, age, native language, presence of a disability;
- Perception of other instructor characteristics such as warmth, approachability, attractiveness;
- Size of class;
- Response rate;
- Grades in class or perception of grades by students;
- Perceived difficulty of class; and,
- Class workload and subject of class, e.g., math, humanities.

Indeed, when the Task Force members reviewed information that other tools such as peer review were also subject to bias, SAI data appeared useful. Task Force members agreed to recommend keeping the SAI, albeit in a revised form, and decreasing SAI data influence in high stakes decisions.

Because of the concerns expressed in the findings about faculty perceptions about perceived bias and preference for formative as opposed to summative use of SAI, some faculty members wanted to “own the data” so that they controlled access and use of the data. If data were to be used primarily as a formative assessment tool to improve instruction, using SAI data summatively in high-stakes decisions such as faculty evaluations and promotion, tenure, and reappointment, they argued, should be discontinued or deemphasized. Consequently, the Task Force recommendations included less use of the data including discontinuation of a table showing average ratings for faculty members to use in dossiers.

Concerns about SAI data ownership extended to student responses to the two open-ended questions. Instructors strongly asserted that responses disproportionately biased readers because these narrative responses are sometimes inaccurate and unverifiable. A single negative comment could seem to carry more weight than positive responses from 30 students to the Likert items even when the comment could not be confirmed by other data. Many faculty strongly disagreed with the practice of allowing department heads access to open-ended responses.

Although faculty member responses to the campus survey were strongly in favor of department head access to SAI data and most department heads and deans stated that reviewing the responses would allow for investigation and follow-up where warranted, strong opposition by some faculty members was voiced to allowing department head access. After much discussion, Task Force members did not choose to recommend a change in the practice of allowing department heads access to open-ended question responses even though some department heads and deans acknowledged problems with open-ended responses.

Recommendations Regarding the Use of Student Assessment Instruments

The Task Force made ten recommendations to the WCU Faculty Senate consistent with the original charge.

1. Students should have the opportunity to provide input on the learning experience including individual courses;
2. Student assessment of courses should include data to be used formatively and summatively. Formative information such as whether the course effectively facilitated achievement of learning outcomes and perceived effort and performance of students should be collected separately, if feasible, from data to

be used for summative purposes, such as whether the instructor meets class and provides feedback on student performance;

3. Student assessment instruments should collect data primarily for formative purposes;
4. Departments should remove reference to specific SAI scores from their collegial review documents and should not assign decision-rules for evaluation to specific SAI scores for annual faculty evaluations and PTR decisions;
5. The Office of the Provost should remove the "SAI table" from the Guidelines for the preparations of dossiers (This recommendation related to a document that showed SAI average ratings. Removing it de-emphasized that practice.);
6. The current instrument for collecting student assessment information should be revised to focus primarily on the learning experience. (See Appendix for suggested questions);
7. Student assessment instruments should allow for the addition of a limited number of questions by programs or departments based on relevant factors such as level of course, content, delivery, length, and others that are specified and approved in the program collegial review document for formative use by the instructor and the program;
8. There should be a structure and process to review the SAI process, including how the data are used on a regular schedule. The first review should occur no later than 2023-2024;
9. Department Heads should participate in mandatory professional development and guidance about how best to assess and incorporate student evaluation data in the annual review processes, including in Collegial Review Committee deliberations and peer review of teaching; and,
10. The Task Force makes no recommendation regarding a vendor or product to employ to collect student assessment information.

Impact

The Task Force submitted an 18-page report to the WCU Faculty Senate in April 2019 accompanied by a document of equal length describing services offered by the Coulter Faculty Commons, the WCU faculty development office, on formative assessment instruments and services. The Faculty Senate and Office of the Provost began immediately to implement recommendations and directed programs and departments to review documents governing annual faculty evaluations and tenure, promotion, and reappointments to eliminate reference to overall SAI averages. Furthermore, the Faculty Senate created committees to review and develop new processes for the student evaluations of instruction.

The Task Force recommendations continue to result in changes to change the SAI instrument and to use the data more appropriately. The Task Force activities informed the Task Force members well about the research on student evaluations and the capabilities of commercial software. Informing others beyond the Task Force and Faculty Senators may not have been successful because it was not a specific charge of the Task Force and not discussed extensively. Task Force members, particularly in the first half of the academic year, listened and sought perceptions of stakeholders in various ways, e.g., forums, informal groups, surveys, but did not endeavor to share what members learned beyond the Faculty Senate

References

- Barre, E. (2015). *Rice CTE student ratings of instruction literature review*. YouTube. https://www.youtube.com/watch?v=DEOruot_gcM&feature=youtu.be
- Barre, E. (2018). *Research on student ratings continues to evolve. We should, too*. Rice University Center for Teaching Excellence. <https://cte.rice.edu/blogarchive/2018/2/20/studentratingsupdate>
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41.
- Doerer, K. (2019, January 18). Colleges Are Getting Smarter About Student Evaluations. Here's How. *Chronicle of Higher Education*. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=134866293&site=ehost-live&scope=site>
- Goos, M., & Salomons, A. (2017). Measuring teaching quality in higher education: Assessing selection bias in course evaluations. *Research in Higher Education*, 58(4), 341-364. <http://dx.doi.org/10.1007/s11162-016-9429->
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94–106.
- Mengel, F., Sauermann, J., & Zolitz, U. (2018). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535-566.
- Mitchell, K., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648-652. 10.1017/S104909651800001X
- Uttl, B., White, C.A., & Gonzalez, D.W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42.

Appendix

Possible Questions for Revised SAI

My Efforts in the Course

1. I attended class:
 - a. Most classes – three or less absences.
 - b. Missed more than three classes.
2. I completed:
 - a. All assignments
 - b. Most assignments
 - c. Few or no assignments
3. The grade I expect in this course is:
 - a. A
 - b. B
 - c. C
 - d. D or F
4. My effort has been:
 - a. Great
 - b. Good
 - c. Average/Okay
 - d. Not good

My Ratings for the Course

Scale: 1 – Strongly Disagree; 2 – Disagree; 3 – Agree; 4 – Strongly Agree

1. This course meets my expectations.
2. I learned most or all the learning outcomes for this course.
3. The instructor communicates goals, objectives, and/or learning outcomes.
4. The instructor communicates expected assignments and due dates.
5. The instructor fosters a class environment of inclusion, respect and dignity for all students.
6. The instructor regularly meets the class when scheduled.
7. The instructor starts class on time and uses most of the scheduled class period.
8. The instructor is prepared and organized for class meetings.
9. The instructor provides a syllabus for the class online or in paper copy.
10. The instructor returns graded assignments in a reasonable timeline.

Open-ended Questions:

1. Do you have any specific recommendations for improving this course?
2. What are one to three specific things about the course or instructor that especially helped to support student learning?

Suggested sources for questions:

<https://teaching.berkeley.edu/course-evaluations-question-bank>

<https://assessment.provost.wisc.edu/best-practices-and-sample-questions-for-course-evaluation-surveys/>

<https://cft.vanderbilt.edu/guides-sub-pages/student-evaluations/>

<https://www.ocadu.ca/Assets/content/teaching-learning/FCDC/Creating+Open-Ended+Questions+for+Student+Feedback+on+Courses+Surveys.pdf>

<http://dept-wp.nmsu.edu/coehpdr/files/2013/07/mid-term-end-term-student-eval.pdf>