UCF | Faculty Senate

Faculty Senate

Agenda for meeting of February 15, 2024, 3:00 – 5:00 p.m.

Location: In person at the Charge on Chamber, Student Union, Room 340
For those unable to make the in person meeting due to travel, location, or health issues, there is a Zoom option:
https://ucf.zoom.us/j/94984421190?pwd=NkFwNXBFV3BXQ1RlaldhWmtjOFBrZz09
Passcode: 541044

1. Call to Order
2. Roll Call via Qualtrics:
3. Approval of Minutes of *January 18, 2024*
4. Recognition of Guests
5. Announcements and Report of the Senate Chair
6. Report of the President
7. Report of the Provost
8. Unfinished Business
9. New Business
    i) Resolution 2023-2024-8 Evaluating Faculty Instruction
10. Committee Reports
    a) B&A Committee: Keri Watson, Chair of B&A Committee
    b) IT Committee: Glenn Martin, Chair of IT Committee
    c) Personnel Committee: Karol Lucken, Chair of Personnel Committee
    d) Research Committee: Linda Walters, Chair of Research Council
    e) Graduate Council: Danny Seigler, Steering Liaison for Graduate Council
    f) Undergraduate Council: Tina Chiarelli, Chair of UCRC, Steering Liaison for Undergraduate Council
11. Campus Climate Report
    a) Undergraduate Teaching Issues
        i) Theodorea Berry, Vice Provost and Dean, College of Undergraduate Studies
12. Other Business
13. Adjournment

Resolution 2023-2024-8: Evaluating Faculty Instruction

Whereas, despite UCF Regulation 3.010 indicating that Student Perceptions of Instruction (SPIs) should not be the only source of evaluating teaching, SPIs remain one of the primary and most convenient methods of evaluating faculty instruction for purposes of annual evaluation, tenure and promotion, and teaching awards at UCF; and

Whereas, empirical research has shown that SPIs are biased against women, with women being judged more harshly than their male counterparts (Boring, 2017; Centra & Gaubatz, 2000; Kogan, Schoenfeld-Tacher, & Hellyer, 2010; Laube, Massoni et al., 2007; Mitchell & Martin, 2018). Empirical research has equally shown that SPIs are biased against ethnic and minority groups, resulting in African American professors being rated, on average, as 21% more mean spirited and 24% harder as compared to Caucasian faculty ratings (Harlow, 2003); and

Whereas, a recommendation of the 2020 report of the UCF SPI Task Force states: "As one of the largest and most innovative universities in the U.S., a designated Hispanic-Serving and Minority Serving institution that is committed to access, inclusion, and diversity, UCF should discontinue the use of SPIs, which perpetuate race- and gender-based biases, in the process of Faculty Performance evaluations" (p.6). The rationale for this recommendation was based in part on an argument that appeared in an issue of Inside Higher Ed, which stated: "Relying on biased instruments to evaluate faculty members is institutional discrimination." (Owen, 2019); and

Whereas, empirical research, including a recent meta-analysis (Uttl, White & Gonzalez, 2017), has shown that SPIs are a poor measure of teaching effectiveness, primarily measuring perceptions of students who are not experts in pedagogy, and are influenced by non-teaching based factors like time of day, subject, and class size (Boring, Ottoboni & Stark, 2016; Flaherty, 2020; Lederman, 2020; Stroebe, 2020); and

Whereas, empirical research has shown that students rate teaching methods that have been proven effective [such as active learning] as less effective than passive learning strategies (Deslauriers, McCarty et al., 2019); and

Whereas, UCF research has shown that less than 60% of students complete SPIs, despite continuous reminders and subsequent barriers to enrollment and other university activities for those failing to complete them (Dziuban, Moskal, Self, & Hubertz, 2022); and

Whereas, UCF research has shown that 66.1% of students from 2017 to 2021 straight lined their SPI responses (Dziuban, Moskal, Self, & Hubertz, 2022); and

Whereas, empirical research has shown that "up to a third of students use instructor ratings to get revenge on instructors they do not like, even to the extent of submitting false information" (Clayson & Haley, 2011; as cited in UCF SPI Task Force Report, 2020:7).

Whereas, empirical research has shown that student grade satisfaction, receiving expected grades, perceived and actual grading leniency, and/or "consumer satisfaction" are important drivers of [positive] faculty evaluations (Johnson, 2002; Eizler, 2002; Felton et al., 2008; Braga et al., 2014; Stroebe, 2020); and

Whereas, empirical research has shown that SPIs, especially when used in high-stake personnel decisions, encourage grade inflation (Johnson, 2006; Shouping, 2005), ultimately affecting the credibility of institutions and creating dubious impressions of student learning and teaching effectiveness; and

Whereas, at UCF, from 2018 to 2023, in lower-level undergraduate courses, 46.8 percent [range of 42.3 - 49] of grades were A's (A /A-) and 26.2 percent [range of 25.3 – 28.2] were B's (B+/B/B-). From 2018 to 2023, in upper-level undergraduate courses, 47.2 percent [range of 44 – 48.9] of grades were A's and 26.1 percent [range of 25.7 - 27.9] were B's (Source:IKM); and

Whereas, at UCF, from 2018 to 2023, the average percentage of A's received in upper-level undergraduate courses was at or exceeded 55 percent [range of 55 – 65] in 6 of 10 colleges. In the remaining 4 colleges, which are responsible for 62% of all grades at UCF, the most commonly reported percentage of A's for upper-level undergraduate courses was 45 percent [range of 31 – 46] and 26 and 36 percent for B's (Data Source: IKM; College of Medicine and Graduate Studies, and Honor's College, where 80 percent of grades are "S," are not included in these figures).

Whereas, research by scholars from Brigham Young, Purdue, and Stanford University (Denning, Eide, Mumford, Patterson & Warnick, 2023) found that the "no direct cost to the university" practice of grade inflation [not changing enrollment patterns, better performance on standardized tests, student-to-faculty ratios or instructional expenditures] is most responsible for increased graduation rates ("The Grade Inflation Conversation We're Not Having .....," April 13, 2023 issue of Chronicle of Higher Education); and

Whereas, four other universities (Colorado-Boulder, Southern California, Oregon, and Kansas) have made substantial changes to the evaluation of faculty teaching, which includes elimination of SPIs as a primary source of evaluating teaching (UCF SPI Task Force, 2020:8-9)

Be it Resolved that UCF abandon use of SPIs in faculty annual evaluations, promotion and tenure, and awards, and require committees, unit/department heads, deans, and other university personnel to employ more objective measures of *teaching quality and commitment* in assessing faculty instruction. Examples of alternative measures include, but are not limited to:

- quality course designations from IDL
- use of evidence-based practices or innovative or FCTL recommended teaching strategies
- creation of new courses for department curriculum
- syllabi, classroom assignments, exams
- grade distributions
- students supervised on independent studies/theses/dissertations
- publications, presentations and/or research with students
- In-class peer observation

Be it Further Resolved that UCF retain use of SPIs for faculty members' personal use in guiding their instruction and in post-tenure review, which complies with current BOG regulations and policies.

Amended Resolution 2023-2024-8: Evaluating Faculty Instruction

Whereas, despite UCF Regulation 3.010 indicating that Student Perceptions of Instruction (SPIs) should not be the only source of evaluating teaching, SPIs remain one of the primary and most convenient methods of evaluating faculty instruction for purposes of annual evaluation, tenure and promotion, and teaching awards at UCF; and

Whereas, empirical research has shown that SPIs are biased against women, with women being judged more harshly than their male counterparts (Boring, 2017; Centra & Gaubatz, 2000; Kogan, Schoenfeld-Tacher, & Hellyer, 2010; Laube, Massoni et al., 2007; Mitchell & Martin, 2018). Empirical research has equally shown that SPIs are biased against ethnic and minority groups, resulting in African American professors being rated, on average, as 21% more mean spirited and 24% harder as compared to Caucasian faculty ratings (Harlow, 2003); and

Whereas, a recommendation of the 2020 report of the UCF SPI Task Force states: "As one of the largest and most innovative universities in the U.S., a designated Hispanic-Serving and Minority Serving institution that is committed to access, inclusion, and diversity, UCF should discontinue the use of SPIs, which perpetuate race- and gender-based biases, in the process of Faculty Performance evaluations" (p.6). The rationale for this recommendation was based in part on an argument that appeared in an issue of Inside Higher Ed, which stated: "Relying on biased instruments to evaluate faculty members is institutional discrimination." (Owen, 2019); and

Whereas, empirical research, including a recent meta-analysis (Uttl, White & Gonzalez, 2017), has shown that SPIs are a poor measure of teaching effectiveness, primarily measuring perceptions of students who are not experts in pedagogy, and are influenced by non-teaching based factors like time of day, subject, and class size (Boring, Ottoboni & Stark, 2016; Flaherty, 2020; Lederman, 2020; Stroebe, 2020); and

Whereas, empirical research has shown that students rate teaching methods that have been proven effective [such as active learning] as less effective than passive learning strategies (Deslauriers, McCarty et al., 2019); and

Whereas, UCF research has shown that less than 60% of students complete SPIs, despite continuous reminders and subsequent barriers to enrollment and other university activities for those failing to complete them (Dziuban, Moskal, Self, & Hubertz, 2022); and

Whereas, UCF research has shown that 66.1% of students from 2017 to 2021 straight lined their SPI responses (Dziuban, Moskal, Self, & Hubertz, 2022); and

Whereas, empirical research has shown that "up to a third of students use instructor ratings to get revenge on instructors they do not like, even to the extent of submitting false information" (Clayson & Haley, 2011; as cited in UCF SPI Task Force Report, 2020:7).

Whereas, empirical research has shown that student grade satisfaction, receiving expected grades, perceived and actual grading leniency, and/or "consumer satisfaction" are important drivers of [positive] faculty evaluations (Johnson, 2002; Eizler, 2002; Felton et al., 2008; Braga et al., 2014; Stroebe, 2020); and

Whereas, empirical research has shown that SPIs, especially when used in high-stake personnel decisions, encourage grade inflation (Johnson, 2006; Shouping, 2005), ultimately affecting the credibility of institutions and creating dubious impressions of student learning and teaching effectiveness; and

Whereas, at UCF, from 2018 to 2023, in lower-level undergraduate courses, 46.8 percent [range of 42.3 - 49] of grades were A's (A /A-) and 26.2 percent [range of 25.3 – 28.2] were B's (B+/B/B-). From 2018 to 2023, in upper-level undergraduate courses, 47.2 percent [range of 44 – 48.9] of grades were A's and 26.1 percent [range of 25.7 - 27.9] were B's (Source:IKM); and

Whereas, at UCF, from 2018 to 2023, the average percentage of A's received in upper-level undergraduate courses was at or exceeded 55 percent [range of 55 – 65] in 6 of 10 colleges. In the remaining 4 colleges, which are responsible for 62% of all grades at UCF, the most commonly reported percentage of A's for upper-level undergraduate courses was 45 percent [range of 31 – 46] and 26 and 36 percent for B's (Data Source: IKM; College of Medicine and Graduate Studies, and Honor's College, where 80 percent of grades are "S," are not included in these figures).

Whereas, research by scholars from Brigham Young, Purdue, and Stanford University (Denning, Eide, Mumford, Patterson & Warnick, 2023) found that the "no direct cost to the university" practice of grade inflation [not changing enrollment patterns, better performance on standardized tests, student-to-faculty ratios or instructional expenditures] is most responsible for increased graduation rates ("The Grade Inflation Conversation We're Not Having .....," April 13, 2023 issue of Chronicle of Higher Education); and

Whereas, four other universities (Colorado-Boulder, Southern California, Oregon, and Kansas) have made substantial changes to the evaluation of faculty teaching, which includes elimination of SPIs as a primary source of evaluating teaching (UCF SPI Task Force, 2020:8-9)

Whereas, there is no single mechanism that effectively evaluates teaching and teaching effectiveness for all UCF faculty, yet many of our teaching evaluations rely solely or predominantly on SPIs, and the complete removal of SPIs from evaluations of teaching and teaching effectiveness would remove the lone mechanism for student input into their courses and instructors, would be incompatible with accreditation standards, and would be in direct conflict with the current UFF-UCF BOT Collective Bargaining Agreement; therefore

Be it Resolved that UCF limit the use of SPIs to no more than one of at least four separate measures used from the list below to evaluate teaching in any faculty annual evaluations, promotion and tenure decisions, post-tenure review, accreditation assessments, and awards, and require committees, unit/department heads, deans, and other university personnel to employ at least four measures of teaching quality and commitment in assessing faculty instruction. Examples of measures include, but are not limited to:

- student perception of instruction
- quality course designations from IDL
- use of evidence-based practices or innovative or FCTL recommended teaching strategies
- creation of new courses for department curriculum
- syllabi, classroom assignments, exams
- grade distributions
- students supervised on independent studies/theses/dissertations
- publications, presentations and/or research with students
- in-class peer observation

Information about Universities and their use of SPIs.

UC Boulder stopped using SPIs and replaced them with a “Faculty Course Questionnaire” and requires multiple means of evaluating faculty:

From UC Boulder Policy APS #1009:

“Faculty members, other than lecturers, teaching regular courses (excluding independent study, thesis, dissertation credits, and similar individual student courses) shall have their overall teaching evaluated annually using multiple means, including normed student feedback on each regular course, addressing behaviors and practices of which students have direct knowledge, using an instrument that attempts to mitigate potential bias in student evaluations of teaching.”

The University of Kansas updated and still uses student evaluation of faculty teaching:

From Office of Provost: Student Evaluation of Teachers, Procedures for Administration of

“The University of Kansas and the department are committed to effective teaching. Students assist in maintaining and enhancing this effectiveness by completing teaching evaluations in a thoughtful and honest manner. We ask that you take time to respond to all questions and write comments. If you do not wish to participate in the process, write "no comment" on the form and turn it in.

Please rate the quality of the course and make suggestions. Student evaluations can help improve instruction and are used in the annual faculty evaluation, reappointment, and promotion and tenure processes.”

The University of Oregon has significantly updated its teaching evaluation process to include peers, students, and the faculty themselves:

From University of Oregon Office of the Provost: Teaching Support and Innovation

“Members of the UO community will be familiar with research on bias inflecting teaching evaluation, particularly the student evaluation of teaching. To mitigate bias and ensure that evaluation supports the development of UO’s teaching culture, the University of Oregon, led by the University Senate and Provost’s office, has launched a new Continuous Improvement and Evaluation of Teaching System.

UO's framework for the Continuous Improvement and Evaluation of Teaching aims to more transparently evaluate teaching with 1) a clear definition of teaching quality and 2) evidence from three sources: peers, students, and faculty themselves. The system also captures and honors the process of teaching improvement, core to the ethos of the Teaching Engagement Program.”

The University of Southern California removed student evaluation of teaching, but at that time they were replaced by a University-wide process for peer-review evaluation of teaching.

From USC Center for Excellence in Teaching

“The following tools and templates are offered in support of schools and departments developing their own peer-review systems. Use of these specific materials is not required by the USC administration, but are intended as examples of best practices in teaching. Schools and departments may choose to adopt the following tools as written, edit them (add, subtract, or reword criteria), or create their own.”

Student Perception of Instruction (SPoI) Task Force Report 2020

AUGUST 2020

UCF Student Perception of Instruction (SPoI) Task Force Report to the Faculty Senate

Taskforce Members

Kelly Allred
Associate Professor, Nursing Practice
Zhongzhu Chen
Assistant Professor, Department of Physics
Lucretia Cooney
Director, Faculty Excellence
Jana Jasinski
Professor, Faculty Excellence
Tamra Legron-Rodriguez
Lecturer, Department of Chemistry
Eric Main
Associate Director, Faculty Center for Teaching & Learning
Ann Miller
Interim Director, Faculty Center for Teaching & Learning
Professor, Nicholson School of Communication and Media
Patsy Moskal
Director, Learning Resources
William Self
Professor and Associate Director of Undergraduate Affairs
Julie Sharek
Instructor, Integrated Business
Keri Watson
Associate Professor, School of Visual Arts & Design
Grace White (Chair)
Associate Lecturer, Department of Psychology

Introduction and Overview

The Student Perception of Instruction (SPoI) Task Force was convened in the spring 2020 in response to Faculty Senate Resolution 2018-19-12 (see http://facultysenate.ucf.edu/resolutions/2018_2019/index.asp) which was focused on improving the Student Perceptions of Instruction at UCF. The work of the task force was focused on reviewing questions and question validity, recommending better methods to evaluate teaching, and defining the role of the SPoI in the evaluation process.

For decades, use of student evaluations in faculty performance has been a hot topic in higher education and a point of significant criticism (e.g. Esarey & Valdez, 2020; Rosen, 2018). The most prominent area of discourse has been related to whether these surveys could reliably and accurately measure teaching effectiveness (Boring, Ottoboni, & Stark, 2016; Emery, Kramer, & Tian, 2003). Moreover, examinations of student evaluations of faculty performance at colleges and universities across the nation have shown a consistent and replicable pattern of bias against female faculty and faculty of color (Boring, 2017; Centra & Gaubatz, 2000; Harlow, 2003; Kogan, Schoenfeld-Tacher, &Hellyer, 2010; Laube, Massoni, Sprague, &Ferber, 2007; McPherson, Jewell, & Kim, 2009; Spooren, Brockx, & Mortelmans, 2013). Thus, the almost exclusive reliance on this biased metric in "high-stakes" personnel decisions like promotion, tenure, and awards can create and perpetuate systemic deficits for faculty who are not white and/or who are female. Therefore, it is incumbent upon universities to consider and weigh the impact of such reliance given the mounting evidence against their validity (Flaherty, 2020; Lederman, 2020).

A Brief History of UCF's SPoI Survey

With regards to the history of SPoI at UCF, there is a body of literature and research on the 16-item SPoI (see appendix 1) which was in use between 1996 and Spring 2013. A brief summary of those studies is presented below:

- Research completed by Wang, Dzuiban, Cook and Moskal (2009) was able to generate general rules to discriminate between faculty rated as excellent and those rated as poor from SPoI data collected from student responses in academic years 1996 to 2001. These findings had practical applications in allowing faculty to be able to target specific areas of student perceptions which in turn may have increased overall ratings.
- Dzuiban, Moskal, Kramer, & Thompson (2012) explored whether there was a difference in the number of elements by which students evaluate their online courses depending on the degree of ambivalence they express about those courses. Further, if there was a difference, what were the dimensions and how did they relate to each other. This research examined student responses in academic years 2008-2010 at UCF. Overall, these data suggested that ambivalence (as indicated by 2, 3, 4 rating on Likert scale) was indicative of a more complex model of student satisfaction (Dzuiban et al., 2012). While students with no ambivalence used a general opinion which determined their evaluations, ambivalent students used multiple categories of information to formulate their ratings. For the most ambivalent (3), the students also evaluate the degree to which the instructor is responsive to them.
- Dzuiban and Moskal (2011) investigated whether the identical student rating instrument is measuring the same or different underlying teaching and learning constructs,

depending on the modality in which the course is offered? Approximately 1.1 million student responses to the 16 item SPI across 3 course modalities (online, blended, face-to-face) at UCF were analyzed. Findings showed data set characteristics for the 3 modalities resulted in a single factor which accounted for 70% of the total system variance. (Dzuiban & Moskal, 2011). The conclusion is: course modality had no impact on students when evaluating educational experiences. Thus, it seems that the same underlying criterion is being used by students when making these ratings regardless of modality.

Based on the above empirical literature, UCF's 16 item assessment appeared to be a valid and reliable measure of **students' perceptions** of instruction. None of the studies examined the current 9-item assessment (see appendix 2) which went into effect in fall 2013. It is possible to infer that the 9 items, which were derived from the 16-item assessment, likely share the same or similar reliability and validity. However, this is an empirical question. The process of assessment validation requires the exploration of the psychometric properties of measures cross-sectionally and over time. These metrics include, and are not limited to, predictive validity, construct validity, and criterion-related validity. Given the lack of this type of assessment validation data for the 9-item measure, the conclusions which can be drawn about its validity are limited. For the 16-item survey, based on the empirical literature, there is still a question of the impact of bias (based on age, gender, race and/or national origin) on these ratings. Specifically, none of the above research addresses or excludes the possibility of bias in the ratings. From a review of the empirical literature, we have a clear picture of how students at UCF viewed excellent and poor instructors (particularly for years 1996-2001). However, whether these subjective assessments are indicative of **objective teaching effectiveness and learning outcomes was not explored.** This question is not answered by the above research.

Given the information provided by the empirical data on the SPoI, we were able to come to a few conclusions and concerns. UCF's 16-item measure seems to have reliably captured **students' perceptions**. However, the question remains as to whether this is a sufficient and functional measure of objective teaching effectiveness and how these ratings connect to learning outcomes. Is teaching effectiveness only to be defined by student perceptions? This research also leaves concerns about bias in the ratings, which is not addressed by the previously reviewed literature. Without more data specific to measurement validation of the 9-item survey, evidence -based conclusions about its validity cannot be made. Further examination of the scope of application, and appropriateness of application, of SPoI as it relates to the objective evaluation of teaching effectiveness should be addressed.

Overview of Task Force Recommendations

This task force's charge is an important and timely one. The national discourse about student surveys highlights the delicate balance between giving students a voice in the academic process and creating an inclusive campus environment for female faculty and faculty of color. Faculty on the task force have also expressed concerns about an inability to introduce effective pedagogy which challenges students' thinking for fear of student reprisals in the evaluation process. Furthermore, concerns about the responsibility for faculty to remind students to complete these surveys, rather than having a university-controlled reminder mechanism, raises additional bias

possibilities. Given the complexity of the problem, the wide variety of concerns, and the high stakes associated with the SPoI, the task force did not arrive at a single, uniform conclusion, but instead came up with the following three possible recommendations for improving the SPoI, or improving the evaluation of teaching and learning in general, from which the Faculty Senate should select:

Recommendation A: Eliminate the Use of Student Perception of Instruction Assessment in Faculty Evaluation.
Recommendation B: Keep SPoI assessment with inclusion of bias awareness language and add additional measures of teaching effectiveness, including instructor reflection, peer review.
Recommendation C: Keep SPoI assessment with edits/changes to items and instructions to increase validity and reduce bias (e.g. bias disclaimers, example prompts)

A detailed summary and report related to each of the above recommendations is included in the next section.

Recommendation A: Eliminate the Use of Student Perception of Instruction Assessment in Faculty Evaluation.

- **Task Force Work Group Members:**

  Tamra Legron-Rodriguez
  Julie Sharek
  Keri Watson

Explanation and Rationale:

In response to a growing body of research, organizations from the American Sociological Association to the Association of American Universities to the American Association of University Professors have issued statements questioning the validity of Student Perceptions of Instruction/Student Evaluations of Teaching, and colleges and universities from the University of Oregon to the University of Southern California have discontinued their use.

Research has demonstrated that SPOIs are:

- Only weakly related to teaching effectiveness
- Used in statistically problematic ways
- Are influenced by factors such as times of day and class size
- Are biased against women, people of color, and adjuncts

As a 2019 American Sociological Association report wrote:

> "Despite the ubiquity of SETs, a growing body of evidence suggests that their use in personnel decisions is problematic. SETs are weakly related to other measures of teaching effectiveness and student learning (Boring, Ottoboni, and Stark 2016; Uttl, White, and Gonzalez 2017); they are used in statistically problematic ways (e.g., categorical measures are treated as interval, response rates are ignored, small differences are given undue weight, and distributions are not reported) (Boysen 2015; Stark and Freishtat 2014); and they can be influenced by course characteristics like time of day, subject, class size, and whether the course is required, all of which are unrelated to teaching effectiveness. In addition, in both observational studies and experiments, SETs have been found to be biased against women and people of color (for recent reviews of the literature, see Basow and Martin 2012 and Spooren, Brockx, and Mortelmans 2015)."

Moreover, as argued in a recent issue of Inside Higher Ed:

> "Relying on biased instruments to evaluate faculty members is institutional discrimination. Indeed, it is simply a matter of time before a class-action lawsuit is filed against an institution for knowingly using biased instruments in evaluating its faculty."

Mechanism for Adoption of Recommendation:

As one of the largest and most innovative universities in the U.S., a designated Hispanic-Serving and Minority Serving institution that is committed to access, inclusion, and diversity, UCF should discontinue the use of SPOIs, which perpetuate race- and gender-based biases, in the process of Faculty Performance evaluations.

Recommendation B: Keep SPoI assessment with inclusion of bias awareness language and add additional measures of teaching effectiveness, including instructor reflection, peer review.

- **Task Force Work Group Members:**
  Eric Main
  Ann Miller
  William Self

Explanation and Rationale:

The idea that student evaluations can stand in for learning rests on two assumptions: that students can accurately gauge their own level of learning, and that they will accurately report those perceptions (Braga et al., 2014). However, a large body of research has shown that students are not good at assessing their own learning (Carpenter et al., 2020; Weinberg et al., 2009). Both laboratory and classroom data demonstrate that students are overconfident in their abilities in comparison to their actual performance (see review by Finn & Tauber, 2014). Many faculty members experience the results of this overconfidence firsthand when students come to them shocked by their poor performance on the first test, even though they thought they understood the material (Carpenter, et al., 2020; McGuire, 2015).

Furthermore, students have a strong tendency to erroneously misinterpret smooth, fluent learning experiences, such as enthusiastic lectures, for learning itself, even though empirical research has shown these teaching approaches to be ineffective or even counterproductive for actual learning (Motz, de Leeuw, Carvalho, Liang, & Goldstone, 2017; Williams & Ceci, 1997). In contrast, retrieval practice, spaced practice, and active learning have been demonstrated to be highly effective (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), but students often state that they do not learn well from these techniques, likely because they are less comfortable with the active nature of the learning experience. Deslauriers, McCarty, Miller, Callaghan, & Kestin (2019) found students rated instructors who employed active learning strategies less highly than those who used passive techniques, even though in reality students taught via active learning methods scored 10% higher on tests over the material. The researchers surmised that active learning by definition involves struggle for students, and students may interpret learning experiences that involve this kind of cognitive exertion as a sign that they are not learning.

As noted in the introductory material, students also seem not to be able to disentangle irrelevant factors from teaching evaluations (Yunker & Yunker, 2003). Research has demonstrated SETs to be affected by personal characteristics such as faculty gender (Weinberg et al., 2009), age (Sprinkle, 2008), nationality (Weinberg et al., 2009), and "hotness" (Felton, Koper, Mitchell, & Stinson, 2004), as well as situational factors like type of course (Uttl & Smibert, 2017), weather at the time of the SET (Braga et al., 2014), and even whether an independent administrator gave students chocolate before they filled out the evaluations (Youmans & Jee, 2007).

Finally, a few studies have found students' accuracy and honesty in reporting to be faulty (Nilson, 2013). For example, a majority of students voluntarily evaluated guest lecturers in their undergraduate and medical school classes who had never taught them (Reynolds, 1977; Uijtdechaage & O'Neal, 2014), and marked their instructor down on promptness of returning assignments even though the instructor had returned all assignments during the entire semester on the following class day (Stanfel, 1995). More disturbing, up to a third of students use instructor ratings to get revenge on instructors they do not like, even to the extent of submitting false information (Clayson & Haley, 2011).

At best, then, student evaluations of teaching (SETs) measure perceived learning, which has little if any relationship to actual learning. However, it is likely that they measure something more akin to satisfaction with the learning experience (Nilson, 2013). Students may not have the same values about teaching that college administrators do. Administrators are concerned that students learn, knowing that learning will make it more likely that they will graduate on time and, subsequently, find employment in their field of study. Students, in contrast, may care primarily about their grades, and secondarily about experiencing a stimulating classroom environment, what Braga and colleagues (2014) refer to as professors' realized utility to students. This sort of satisfaction may be the basis of SET scores, scores that are commonly misinterpreted by administrators as teaching effectiveness.

Nevertheless, quantitative end-of-semester SETs are the most commonly used technique for assessing the quality of teaching among college faculty. Up to 94% of deans and administrators use them to inform a variety of personnel decisions (Miller & Seldin, 2014). Typically administered as electronic surveys with Likert-type items inquiring into teachers' clarity, organization, and caring for students, SETs are likely pervasive because they are time efficient and inexpensive to administer. But it also makes intuitive sense that students, who are in the position to directly observe both their own learning and their instructors' teaching, should have a major voice in providing input about their classes. This option, therefore, does not propose to do away with student evaluations, but to supplement them with other measures, so as to triangulate evaluation of teaching quality.

Proposed Improvement(s):

The AAUP (2015) *Statement on Teaching Evaluation* states that firsthand data from various sources should be gathered, including from students, but emphasizes the primacy of faculty colleague judgements of teaching effectiveness. They suggest that the following types of data should be systematically gathered: 1) factual description of what an individual does as a teacher including number and level and kinds of classes taught, the numbers of students, out-of-class activities related to teaching, course syllabi, tests, materials, and methods employed in instruction; 2) various measures of the effectiveness of these efforts including data from students, trained observers, faculty colleagues, and self-evaluation; and fair consideration of the relation between these efforts and expectations of the department and institution. (Additional ideas related to STEM education can be found at https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/P%26T-Matrix.pdf.)

In this regard, we identified four institutions that can serve as aspirational models for UCF if we are to move toward a multi-measure evaluation of faculty teaching: the University of Southern California, the University of Oregon, the University of Kansas, and the University of Colorado at Boulder,

University of Southern California has developed a detailed peer review system. The shift was featured in an article in *Chronicle of Higher Education* (Supiano, 2018). Detailed resource for peer review are available at the USC Center for Excellence in Teaching website.

University of Oregon has developed a holistic framework for teaching assessment than include peer review, self-reflection and student feedback. Detailed information is available about their procedures on the Provost's web page.

University of Kansas is currently in the midst of a 5-year National Science Foundation grant to develop a framework called Benchmarks for Teaching Effectiveness. The framework includes evaluation of teaching in seven areas, one of which is student perceptions. The university encourages the synthesis of information from instructor, peers, and students in departmental and school level evaluation.

Also funded by an NSF grant, the Teaching Quality Framework at the ***University of Colorado Boulder*** draws from multiple source of evidence to evaluate teaching include "voices" provided by the instructor, peer feedback, and student voices. The framework defines teaching as a scholarly activity with seven core elements.

The following principles are shared by all four of these institutional efforts:

- Student evaluations of teaching should not be the sole evidence on which teaching effectiveness is judged.
- At a minimum, three inputs should be included in the evaluation of teaching system: student feedback, instructor reflection, and peer review.
- Evaluation should be tailored by departments to make it appropriate to the discipline.
- Some shifts of wording are required in student evaluation instruments to make them appropriate for student input. The focus of these changes varies by institution.

Mechanism for Adoption of Recommendation:

Adoption of this recommendation would entail creation of a multi-disciplinary task force that would investigate peer and self-evaluation instrumentation currently available and pilot selected instruments across a range of departments. Based on the experience of the above-cited institutions, this would need to be a multi-year process in order to receive feedback across a range a disciplines and achieve faculty and departmental buy-in for the final product.

Recommendation C: Keep SPoI assessment with edits/changes to items and instructions to increase validity and reduce bias (e.g. bias disclaimers, example prompts)

- **Task Force Work Group Members:**
  Zhongzhu Chen
  Patsy Moskal
  Grace White

Explanation and Rationale:

Student rating data can give voice to student experiences and concerns in the classroom. Thus, there may be some hesitation to abandon or diminish the student's perspective. However, student perspectives can also be tainted by personal biases unrelated to course content or instructor performance (Esarey & Valdez, 2020). As colleges and universities across the nation grapple with increased scrutiny and criticism of student surveys, many are choosing to try to address and mitigate these issues of bias (Flaherty, 2019; Peterson, Biederman, Andersen, Ditonto, & Roe, 2019). Evidence-based means to diminish these biases are lacking. Nonetheless, one possibility which has some empirical support, in the short-term, is cuing students to be aware of their biases prior to completing surveys (Peterson et al., 2019). This "cuing" is done through a statement or "disclaimer" which students read before making ratings on their instructors. Experimental research has shown that bias disclaimers can improve (or reduce) negative bias by up to .5 of a point for female faculty (Peterson et al., 2019). The improvements found in this research suggest bias disclaimers may be an effective tool in addressing gender bias in the short-term. However, their impact on racial bias and ageism was not explored (Peterson et al., 2019). Additional research which examines the long-term impact of these disclaimers must be completed to fully understand their effectiveness.

Similar to "cuing" students to their bias, it may also be beneficial to provide students with additional instruction and guidance on how to appropriately interpret each item on the evaluation form. Beyond bias, it has been called into question as to whether students have the ability to accurately review instructional quality (Jimaa, 2013). Therefore, an additional criticism of student rating data in faculty performance argues that students generally lack the knowledge, motivation, or perspective on the learning process, to provide meaningful evaluations of teaching. As a result, students' ratings on the items reflect more of their subjective feeling towards the course and the instructor, instead of a more objective judgement of the quality of instruction. The ratings could also be strongly influenced by comparison with other courses that the student happen to be taking simultaneously, resulting in a bias against more rigorous and challenging courses. If this is true, then providing students with assistance in framing their classroom experiences may be of benefit. One such method of providing context is to give examples of specific approaches, strategies, or experiences in the class being evaluated, which would be appropriate for receiving a Poor (1) rating or would be appropriate for receiving an Excellent (5) rating. These "example response prompts" would provide students with a mental framework, or context, upon which to gauge the students' experiences in the class and base course ratings. Thus, these prompts can steer students to think of specific types of relevant information when evaluating said experiences.

Proposed Improvement(s):

Bias Disclaimer. If student rating data are used in faculty evaluation, there must be an attempt to alleviate any impact of bias against under-represented faculty. Thus, we recommend that UCF adopt a bias disclaimer to be included in the SPoI prior to students' completion of their instructor ratings. **An example of an example bias disclaimer which could be incorporated into the SPoI is as follows:**

> Student evaluations of teaching play an important role in the review of faculty. Your opinions influence the review of instructors that takes place every year. University of Central Florida recognizes that student evaluations of teaching are often influenced by students' **unconscious** and **unintentional** biases about the race and gender of the instructor. Women and instructors of color are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned.
>
> As you fill out the course evaluation please keep this in mind and make an effort to resist stereotypes about professors. Focus on your opinions about the content of the course (the assignments, the textbook, the in-class material) and not unrelated matters (the instructor's appearance). (*Adapted from Peterson et al., 2019*)

Due to the lack of long-term data on the effectiveness of these disclaimers, the university must recognize the role which bias may play in student ratings, if they are used for important decisions related to promotion, tenure, and awards. Given that there is no fool-proof method to root out bias, departments and colleges must take into consideration how much weight, or value, should be attributed to these ratings when making such decisions.

Example Prompts. It is important to emphasize that this work group does not recommend that a set of "one size fit all" example description should be imposed on all student evaluation forms, as it will be impossible for find examples that are general to all disciplines, all course sizes, and all models of delivery. Instead, we recommend that faculty and departments should be able to customize these "example response prompts" to their specific domain and/or course content. Thus, a requirement to create an adaptable SPoI assessment method or system, which appropriately addresses differences in learning requirements across content domains, course sizes and delivery methods, is a necessity. It is the consensus of our work group that a problem complicated as evaluation of teaching effectiveness could only be resolved by enabling and encouraging all faculty and administration across campus to engage in active discussion about the definition of "good teaching" in different context. To increase the validity of assessments, the SPoI must be able to address variation in teaching methods and/or modality for the example prompts. The current document provides examples of what those example prompts could look like. In other words, the following list is a "example of example", which we hope could serve as the seed for future conversation on teaching effectiveness. **An example of "example" prompts for SPoI items which could be incorporated into the SPoI are as follows, (see appendix 3 for all items):**

1. **Effectiveness organizing the course**

> **An instructor could receive a "1" rating** for **Effectiveness organizing the course** if for example: for a Webcourse that has no modules, no headings or titles for information, no guidance for navigation, made frequent changes to the course format, assignment deadlines, F2F- content seems jumps from idea to idea, no consistency in presentation, unannounced changes to deadlines, etc.
>
> **An instructor could receive a "5" rating** for **Effectiveness organizing the course** if for example: Most assignments and course materials are accessible via more than one method, with clear instruction on how to access and utilize. Different components of the course, such as homework, exams, lecture and reading materials, are well aligned with each other. Provides reasonable flexibility in schedule for students without compromising the rigor of instruction.

As stated with the "bias disclaimer," long-term evaluations of the effectiveness of this method in improving students' precision in course evaluation would be needed.

Appropriate Application/Use of Ratings. Until the effectiveness of bias reduction using these methods is known, we encourage use of other evidence-based practices in evaluating faculty performance. Given the possible bias inherent to student rating data, we urge departments, colleges, and administrators to the view these data as "feedback" rather than as formal ratings. **There are also several statistical concerns and recommendations which must be taken into account for the appropriate use and application of these ratings.** These statistical issues include: (1)low response rates, (2)class size, (3) use of averages on categorical data, and (4) comparisons between distributions of scores (Stark & Freishtat, 2014).

Any statistician would state that use of metrics from a sample in which the response rate is low cannot be generalized to the larger population. Thus, making inferences about faculty performance if only a small portion of students have responded to their course SPoI may be inappropriate. Similarly, the average SPoI scores in small classes will be more greatly influenced by outliers, luck, and error (Stark & Freishtat, 2014). Therefore, instructors who teach smaller classes may be more affected by student rating bias, given that the *mean* is sensitive to extremes within the dataset. Hence, both low response rate and small class sizes may endanger faculty ratings, making these scores more vulnerable to bias.

It is of note that SPoI responses are ordinal categorical variables in which students make ratings from Poor (1) to Excellent (5). Stark and Freishtat (2014) point out that these student rating numbers are labels, not values. Thus, one cannot assume the difference between one and two is the same as the difference between four and five. Statistically, it does not make sense to average categorical variables. The appropriateness of use of parametric statistics with data which have Likert response formats continues to be debated among those who use statistics in the social and behavioral sciences (Leung, 2011). It is crucial that those making decisions from interpretations of the data understand the categorical nature of the variables and the appropriate ways in which to analyze these data. Lastly, if SPoI averages were statistically meaningful, it is improper to compare them with other scores, such as the departmental average, without knowing the distribution of scores (Stark & Freishtat, 2014). To further this point, it may be inappropriate to compare SPoIs of very different classes. Comparing the average without knowing the distribution, leaves out meaningful and required information for accurate interpretation. As an academic institution, utilization of scientifically rigorous methods to validate, implement, and interpret assessments must be our standard procedure. Overall, we must continue to explore

evidence-based methods to evaluate “effective” teaching, while understanding that these student ratings provide a portion of a larger picture in the totality of faculty performance.

Mechanism for Adoption of Recommendation:

Adoption of this recommendation would require alterations to the content of the current SPoI survey as well as the need for a more adaptive system of assessment. The inclusion of a “bias disclaimer” statement prior to students accessing the survey items would be required. Thus, students must view, read, and agree to proceed in order to complete their course evaluations. With regards to the additional “example prompts,” given the dynamic nature of these prompts depending upon the course or content area the SPoI system should be more dynamic and adaptable to specific courses. This requires the exploration of newer assessment system or technology which allows this type of customization. If this recommendation is selected an additional task-force or committee should be formed with the focus on this task as it relates to the technical aspects of the implementation of the desired changes. Lastly, departments and colleges must use and implement the statistically appropriate procedures for calculating and interpreting these measures. The inappropriate statistical application and comparison of the mean ratings can compound the impact of bias.

Summary of Report

As indicated by the Faculty Senate Resolution 2018-19-12, the university must take action to address its use and application of the Student Perception of Instruction (SPoI) survey in its current form. As an institution of higher learning, we cannot ignore, nor be complacent, about growing concerns and evidence of bias in student rating data. Attempts to mitigate the impact of bias can follow the three possible recommendations for improving the SPoI, or improving the evaluation of teaching and learning in general, which are to:

(a) eliminate the use of SPoI in faculty evaluation,
(b) keep the current SPoI with inclusion of bias awareness language and add additional measures of teaching effectiveness, OR
(c) keep the current SPoI with edit/changes to items and instructions to increase validity and reduce bias.

The university must also encourage a holistic approach to evaluating faculty performance across all departments and colleges at the institution. At its very best, research on student data suggests that these ratings only account for **18% of the variance in how much students learn** (Kornell & Hausman, 2016). Consequently, over reliance on these "simple" metrics, like SPoI averages, can lead to inaccurate and unfair judgments of faculty.

It is of note that multiple professional organizations have urged universities to move away from a primary focus on student ratings in the evaluation of teaching (Flaherty, 2019). The American Sociological Association has been leading the charge in concert with other organizations in support of a cultural and institutional shift away from a reliance on these flawed metrics. As cited in this report, there is a significant body of research which suggests that use of these ratings in important personnel decisions leads to systemic bias against vulnerable groups, particularly women and people of color. Public universities, with similar size and scope as UCF, have been able to implement substantive changes to their faculty review process and student rating procedures the benefit of faculty and students (Flaherty, 2019).

Even more striking is the impact that the sole reliance on student rating data has on faculty pedagogy. Researchers posit that institutions which strongly depend on student rating data foster a culture of decreased rigor in their educational practices (e.g. Stroebe, 2016). Many faculty members across the nation also believe that decreasing educational rigor can increase student ratings (Morgan, Sneed, & Swinney, 2003). Thus, reliance on these ratings may have a counter-productive effect of increasing grade inflation while reducing the quality and impact of teaching. Not only are students spending less time engaged in the academic process, there also appears to be a significant *decrease* in improvements in critical thinking skills among more recent college graduates in comparison to college graduates of previous decades (Arum & Roksa, 2011; Pascarella et al., 2011). Therefore, universities must be willing to devote the time and resources to assess faculty accurately and fairly. Attempts to short-cut this process can only lead to biased and unjust evaluations which primarily hurt women and people of color. Moreover, an unwillingness to invest in a dynamic model of faculty evaluation also hurts the students whom institutions serve. Adoption of one of the recommendations of this task force, as they relate to the application and implementation of the SPoI survey, would better serve UCF's core values of integrity, scholarship, community, and excellence in how we make decisions about faculty performance.

REFERENCES

American Association of University Professors. (2015). Statement on teaching evaluation. Retrieved 19 August, 2020 from https://www.aaup.org/report/statement-teaching-evaluation.

Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences*, *27*(2), 184–201. http://doi.org/10.1177/0739986304273707

Arum, R., & Roksa, J. (2011). Limited learning on college campuses. *Society*, *48*, 203–207.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, *145*, 27-41.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education review*, 41, 71-88.

Brunsma, D. L., Feagin, J. R., & McKinney, K. D. (2003). The Many Costs of Racism. *Contemporary Sociology*. Rowman & Littlefield. http://doi.org/10.2307/1556631

Carpenter, S. K., & Witherby, A. E. (2020). On students' (mis)judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition,* https://doi.org/10.1016/j.jarmac.2019.12.009

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3), 409-432.

Centra, J. A., & Gaubatz, N. B. (2000). Is There Gender Bias in Student Evaluations of Teaching? *The Journal of Higher Education*, *71*(1), 17. Retrieved from http://www.jstor.org/stable/2649280?origin=crossref

Clayson, D. E. (2009). Student Evaluations of Teaching: Are They Related to What Students Learn?

Journal of Marketing Education, *31*(1), 16–30. http://doi.org/10.1177/0273475308324086

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*, 19251-19257.

Dovidio, J. F., Gaertner, S. E., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity & Ethnic Minority Psychology*, *8*(2), 88--102. http://doi.org/10.1037//1099-9809.8.2.88

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nthan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Dziuban, C. D. & Moskal, P. D. (2011). A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments. *Internet and Higher Education*, *14*, 236-241.

Dziuban, C. D., Moskal, P. D., Kramer, L., & Thompson, J. (2013). Student satisfaction with online learning in the presence of ambivalence: Looking for the will-o'-the-wisp. *Internet and Higher Education*, *17*, 1-8.

Esarey, J. & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assessment and Evaluation in Higher Education*, 1-15. doi.org/10.1080/02602938.2020.1724875

Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2004). Web-based student evaluations of professors: The relations between perceived quality, easiness and sexiness. *Assessment & Evaluation in higher education*, *29*, 91-108.

Flaherty, C. (2020, Feb 27). *Study: Student evaluations of teaching are deeply flawed*. Inside Higher

Ed. https://www.insidehighered.com

Flaherty, C. (2019, May 20). *Fighting gender bias in student evaluations of teaching, and tenure's effect on instruction*. Inside Higher Ed. https://www.insidehighered.com

Flaherty, C. (2019, Sep 10). *Sociologists and more than a dozen other professional groups speak out*. Inside Higher Ed. https://www.insidehighered.com

Harlow, R. (2003). "Race Doesn't Matter, but...": The Effect of Race on Professors' Experiences and Emotion Management in the Undergraduate College Classroom. *Social Psychology Quarterly*, *66*(4), 348. http://doi.org/10.2307/1519834

Jimaa, S. (2013). Students' Rating: Is it a measure of an effective teaching or best gauge of learning? *Procedia - Social and Behavioral Sciences*, *83*, 30-34. 10.1016/j.sbspro.2013.06.006.

Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. W. (2010). Student evaluations of teaching: perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, *15*(6), 623–636. http://doi.org/10.1080/13562517.2010.491911

Kornell, N. & Hausman, H. (2016). Do the best teachers get the best ratings? *Frontiers in Psychology*, *7*, 1-8. doi: 10.3389/fpsyg.2016.00570

Laube, H., Massoni, K., Sprague, J., & Ferber, A. (2007). The impact of gender on the evaluation of teaching: What we know and what we can do. *National Women's Studies Association Journal*, *19(3)*, 87-104.

Lakin, A. L. (2016). Effective faculty evaluation at the teaching-centered university. *International Journal of Educational Management*, *30*(6), 976–988. http://doi.org/10.1108/IJEM-03-2015-0030

Lederman, D. (2019, Aug 19). *Many colleges are abandoning or downgrading student evaluations during coronavirus. Will that stick?* Inside Higher Ed. https://www.insidehighered.com

Leung, S. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-Point Likert scales, *Journal of Social Service Research*, *37*(4), 412-421, doi: 10.1080/01488376.2011.580697

Martinez, M. A., & Welton, A. D. (2015). Straddling Cultures, Identities, and Inconsistencies: Voices of Pre-Tenure Faculty of Color in Educational Leadership. *Journal of Research on Leadership Education*, *12*(2), 122–142. http://doi.org/10.1177/1942775115606177

McGuire, S. Y. (2015). Teach students how to learn: Strategies you can incorporate into any course to improve student metacognition, study skills, and motivation. Sterling, Virginia: Stylus.

McPherson, M. A., Jewell, R. T., & Kim, M. (2009). What Determines Student Evaluation Scores? A Random Effects Analysis of Undergraduate Economics Classes. *Eastern Economic Journal*, *35*(1), 37–51. http://doi.org/10.1057/palgrave.eej.9050042

Motz, B. A., de Leeuw, J. R., Carvalho, P. F., Liang, K. L., & Goldstone,R. L. (2017). A dissociation between engagement and learning: Enthusiastic instructions fail to reliably improve performance on a memory task. PLoS ONE, 12 (7). Doi: 10.1371/journal.pone.0181775

Nilson, L. (2013). Measuring student learning to document faculty teaching effectiveness. In J. E. Groccia & L. Cruz, (Eds.) To Improve the Academy: Resources for Faculty, Instructional, and Organizational development, vol. 32 (pp. 287-299). Jossey-Bass.

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How robust are the findings of Academically Adrift? *Change: The Magazine of Higher Learning*, *43*, 20–24. https://doi.org/10.1080/00091383.2011.568898

Perry, G., Moore, H., Edwards, C., Acosta, K., & Frey, C. (2008). Maintaining Credibility and Authority

as an Instructor of Color in Diversity-Education Classrooms: A Qualitative Inquiry. *The Journal of Higher Education*, *80*(1), 80–105. http://doi.org/10.1353/jhe.0.0030

Peterson, D., Biederman, L., Andersen, D., Ditonto, T., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLOS ONE*, *14*(5): e0216241. https://doi.org/10.1371/journal.pone.0216241

Reynolds, D. V. (1977). Students who haven't seen a film on sexuality and communication prefer it to a lecture on the history of psychology they haven't heard: Some implications for the university. Teaching of Psychology, 4, 82-83.

Smith, W. A., Allen, W. R., & Danley, L. L. (2007). "Assume the Position . . . You Fit the Description." *American Behavioral Scientist*, *51*(4), 551–578. http://doi.org/10.1177/0002764207307742

Sprinkle, J. E. (2008). Student perceptions of effectiveness: An examination of the influence of student biases. College Student Journal, 42, 276-294.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research*, *83*(4), 598–642. http://doi.org/10.3102/0034654313496870.

Stark, P. & Frieshtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research* (doi: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1)

Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. *Perspectives on Psychological Science*, *11*(*6*), 800-816. doi: 10.1177/1745691616650284.

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, *62*(4), 271–286. http://doi.org/10.1037/0003-066X.62.4.271

Supiano, B. (2018). A university overhauled its course evaluation to get better feedback. Here's what changed. Chronicle of Higher Education, Retrieved 19 August 2020 from https://www.chronicle.com/article/A-University-Overhauled-Its/243803?key=d70YTgktpgcK7_T5bad8tRqILQK6jx5UJw3cLsvnYrvPWD5cqMtK2llNHCSLCrqWdnpEaWk5Q21ScVYxNWFsZTVMZ29qakh5SDlESjUyMnlxV3ZJbkROdWFZbw

Stanfel, L. E. (1995). Measuring the accuracy of student evaluations of teaching. Journal of Instructional Psychology, 22, 117-125.

Tatum, H. E., Schwartz, B. M., Schimmoeller, P. A., & Perry, N. (2013). Classroom participation and student- faculty interactions: Does gender matter? *Journal of Higher Education*, *84*(6), 745–768. http://doi.org/10.1353/jhe.2013.0036

Tuitt, F., Hanna, M., Martinez, L. M., Salazar, C., & Griffin, R. (2009). Teaching in the Line of Fire : Faculty of Color in the Academy. *Thought & Action*, 65–74. Retrieved from http://beta.nsea-nv.org/assets/docs/HE/TA09LineofFire.pdf

Uijtdehaage, S., & O'Neal, C. (2015). A curious case of the phantom professor: Mindless teaching evaluation by medical students. Medical Education, 49, 928-932.

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. http://doi.org/10.1016/j.stueduc.2016.08.007

Wang, M.C., Dziuban, C.D., Cook, I.J., Moskal, P.D. (2009). [Dr. Fox Rocks: Using Data-mining Techniques to Examine Student Ratings of Instruction]. In M. C. Shelley II, L. D. Yore, & B. Hand (Eds.), Quality research in literacy and science education: International perspectives and gold standards. Dordrecht, The Netherlands, Springer, pp383-398.

Weinberg, B. A., Hashimoto, M., & Fleisher, B. M. (2009). Evaluating teaching in higher education.

The Journal of Economic Education, 40(3), 227-261.

Williams, W. M., & Ceci, S. J. (1997). “How’m I doing?” Problems with student ratings of instructors and courses. Change: The Magazine of Higher Learning, 29, 12-23.

Youmans, R. J., & Jee, B. D. (2007). Fudging the numbers: Distributing chocolate influences student evaluations of an undergraduate course. Teaching of Psychology, 34, 245-247.

Yunker, P. J., & Yunker, J. A. (2003). Are student evaluations of teaching valid? Evidence from an analytical business core course. Journal of Education for Business, 78(6), 313-317.

Appendix 1

Student Perception of Instruction Survey (Prior to 2013)

From 1996 to Spring 2013, the SPI process consisted of sixteen multiple choice questions and four free response comment questions.

The multiple choice questions included:

1. Feedback concerning your performance in this course was
2. The instructor 's interest in your learn in g was
3. Use of class time was
4. The instructor 's overall organization of the course was
5. Continuity from one class meeting to the next was
6. The pace of the course was
7. The instructor 's assessment of your progress in the course was
8. The texts and supplemental learning materials used in the course were
9. Description of course objectives and assignments
10. Communication of ideas and inform at ion
11. Expression of expectations for performance
12. Availability to assist students in or outside of class
13. Respect and concern for students
14. Stimulation of interest in the course
15. Facilitation of learning
16. Overall assessment of instructor

Possible responses were Excellent, Very Good, Good, Fair, and Poor.

The free response comment questions included:
17. The thing (s) I like the MOST about this course
18. The thing (s) I like the LEAST about this course
19. What is your reaction to the method of evaluating your mastery of the course (i.e. , testing , grading, out of class assignments (term papers), instructor feed back , et c.)
20. Additional comments and suggestions for improvement

Multiple choice questions 1 through 8, and the comment s, were considered confidential and used only for instructor evaluation. However, the response to questions 9 to 16 were public information published by the university.

Appendix 2

Current Student Perception of Instruction (Spring 2013 to present)

In this version of the SPI, there are currently nine multiple choice questions and two free response comment questions, down from 16 and four prior to Spring 2013. The number of questions was reduced in the hopes of increasing student participation/response rates.

The multiple choice questions are

1. Effectiveness organizing the course
2. Effectiveness explaining course requirements, grading criteria, and expectations
3. Effectiveness communicating ideas and/or information
4. Effectiveness showing respect and concern for students
5. Effectiveness stimulating interest in the course
6. Effectiveness creating an environment that helps students learn
7. Effectiveness giving useful feedback on course performance
8. Effectiveness help in g students achieve course objectives
9. Overall, effectiveness of the instruction

Possible responses were Excellent, Very Good, Good, Fair, and Poor.

The free response comment questions are:

l. What did you like best about the course and/or how the instructor taught it?
2. What suggestions do you have for im proving the course and/or how the instructor taught it?

On November 30, 2013, the Faculty Senate approved the web publication of the responses for all nine multiple choice questions for Spring 2013 onward . However, the comments are still confidential. These SPI files can be found at http://net2865.net.ucf.edu/.

Appendix 3

Example for example SPoI prompts

1. An instructor could receive a "1" rating for **Effectiveness organizing the course** if for example: for a Webcourse that has no modules, no headings or titles for information, no guidance for navigation, made frequent changes to the course format, assignment deadlines, F2F- content seems jumps from idea to idea, no consistency in presentation, unannounced changes to deadlines, etc.
   An instructor could receive a "5" rating for **Effectiveness organizing the course** if for example: Most assignments and course materials are accessible via more than one method, with clear instruction on how to access and utilize. Different components of the course, such as homework, exams, lecture and reading materials, are well aligned with each other. Provides reasonable flexibility in schedule for students without compromising the rigor of instruction.

2. An instructor could receive a "1" rating for **Effectiveness explaining course requirements, grading criteria, and expectations** if for example, for a Webcourse and F2F, does not provide a written explanation of course expectations, does not provide any guidelines or grading criteria

   An instructor could receive a "5" rating for **Effectiveness explaining course requirements, grading criteria, and expectations** if for example,
   Clearly communicated the expectations and grading schemes for the course early on, and remind students frequently during the semester.

3. An instructor could receive a "1" rating for **Effectiveness communicating ideas and/or information** if for example, for a Webcourse and F2F, students cannot understand or follow what the instructor is saying/writing, and the instructor makes little effort to adjust or improve over the semester.
   An instructor could receive a "5" rating for **Effectiveness communicating ideas and/or information** if for example, for a Webcourse and F2F, the instructor utilized multiple methods to communicate idea/information, and students can understand the information with little difficulty.

4. An instructor could receive a "1" rating for **Effectiveness showing respect and concern for students** if for example, for a Webcourse and F2F, never responds to student questions or emails, does not provide feedback on assignments, do not make schedule adjustments for unexpected hardship such as a hurricane.

   An instructor could receive a "5" rating for **Effectiveness showing respect and concern for students** if for example, for a Webcourse and F2F, actively reach out to students about their progress and difficulty, provide useful feedback to students, devotes extra effort to accommodate students with special needs such as providing alternative exam times.

5. An instructor could receive a "1" rating for **Effectiveness stimulating interest in the course** if for example, for a Webcourse and F2F, information discussed in course is un-useful or interesting to that specific topic

An instructor could receive a "5" rating for **Effectiveness stimulating interest in the course** if topically applicable instructional activities are presented in a manner that is motivating and relevant to students.

6. An instructor could receive a "1" rating for **Effectiveness creating an environment that helps students learn** if for example, for a Webcourse and F2F, does not provide an accessible inclusive classroom, (such as students are unable to access course materials/documents/text, etc or not all students have the ability to participate, engage with instructor)
   An instructor could receive a "5" rating for **Effectiveness creating an environment that helps students learn** if all course materials are easily accessible and inclusive for all students and all students have ample opportunities to participate and engage with the instructor.
7. An instructor could receive a "1" rating for **Effectiveness giving useful feedback on course performance** if for example, for a Webcourse and F2F, does not provide information on how to improve on specific course topics, or does not provide corrective instruction on assignments, etc
   An instructor could receive a "5" rating for **Effectiveness giving useful feedback on course performance** if the course provides students with rubrics and/or details on how they can succeed and/or improve on specific course assignments and assessments
8. An instructor could receive a "1" rating for **Effectiveness helping students achieve course objectives** if for example, for a Webcourse and F2F, does not state or provide learning objectives, workload is not enough to engage students or overwhelming; not doable.
   An instructor could receive a "5" rating for **Effectiveness helping students achieve course objectives** if course learning objectives are clearly delineated and the course workload is appropriate.
9. An instructor could receive a "1" rating for **Overall effectiveness of the instructor** if for example, the instructor receives either a 1 or a 2 in all or most of the other categories, and makes little effort to improve the overall quality of the course.
   An instructor could receive a "5" rating for **Overall effectiveness of the instructor** if for example, the instructor receives either a 4 or a 5 in all or most of the other categories, and provided students with an exceptional learning experience while holding academic rigor.

New UCF Academic Calendar

Faculty and Staff
At a Glance

NEW ACADEMIC CALENDAR

What's happening?
UCF is modifying its academic calendar to include accelerated term options. Colleges will determine the availability of course offerings.

Why?
It provides colleges with greater scheduling flexibility and provide students with the flexibility of shorter pathways to earning their degree.

When?
The new academic calendar will be introduced in Summer 2024.

NEW CLASS SCHEDULING OPTIONS

UCF will introduce the new academic calendar in the 2024-25 academic year. The modifications are:

- Three sessions in the fall and spring semesters
  - Session F: First eight weeks
  - Session G: Second eight weeks
  - Session 1: Full 16-week term

- Two accelerated session options:
  - A three-weeks asynchronous Online Winter Intersession that takes place Mid-December through early January
  - Maymester session that spans four weeks in May

- The summer semester will continue to offer four sessions A, B, C, and D

REGISTRATION

2024 Summer and Fall registration begins on March 28

Students will be able to registers for available courses for the following 2024 sessions:

- Maymester: May 6 – 13
- Summer (Session A, B, C, D)
- Fall and Spring
    - Session F: First eight weeks - August 19 – October 12
    - Session G: Second eight weeks - October 14 – December 7
    - Session 1: Full 16-week term - August 19 – December 7
- Asynchronous Online Winter Intersession: December 16, 2024 – January 3, 2025

IMPORTANT DATES

March 11: New academic calendar available at calendar.ucf.edu

May 6 – 13: Maymester

Summer:

Fall Session F: August 19 – October 12

Fall Session G: October 14 – December 7

Fall Session 1: August 19 – December 7

Winter Intercession: December 16, 2024 – January 3, 2025

KEY COMMUNICATION DATES

March 1:
- New Academic Calendar webpage is published.
- Faculty and staff receive advance copy of student communication.

March 4: Students notified of the new calendar.

March 11
- New academic calendar published.
- Enrollment appointment dates available for Summer 2024 and Fall 2024.

March XX
- Registration communications begin and will include new academic calendar messaging.

March 28
- Summer 2024 and Fall 2024 registration begins.

.

Proposal:

Eliminate the C-/D+/D- as grades at the University of Central Florida.

Rationale:

The C- grade causes a tremendous amount of confusion. Many majors require students to earn at least a 2.0 in major courses. This means that a C- equates to an unsuccessful completion of a course.

To add to this confusion, since the +/- system is optional, the current system leads to a situation where students with the exact same average could have vastly different results for their academic career. This is best explained with examples.

Student A takes AMH 4170 in spring 2022. Their professor does not use the +/- system. They earn a 72 in the course and receive a C. The class counts for the History major.

Student B takes AMH 4170 in summer 2022. Their professor does use the +/- system. They earn a 72 and receive a C-. The class does not count for the History major.

Also, imagine if Student A and B were both on probation. Student A would remain on probation while student B would be dismissed from UCF.

In theory, this could even happen if the students took different sections of the same class in the same semester.

Similarly, as we expand the use of the S/U system, the C- (which equates to U) is potentially going to cause more problems.

Eliminating the C- would also help improve our graduation rate (as students are not forced to retake these classes.) While this might be seen by some as "watering down" our requirements, in reality, many professors who award C- do not realize the consequences of the grade. And, we must consider whether the current policy makes sense: does a C- reflect the unsuccessful completion of a course? It is easy to make an argument that instead it reflects the minimally successful completion of a course.

Perhaps because of these complications, the C- is rarely used. In Fall 2021, only .51 percent of all grades at UCF were C-. (D- and D+ are used even less--.14 and .49 percent). So, in total this change only impacts 1.14 percent of the grades given at UCF.

Note: At least one of our fellow SUS institutions (FIU) does not use the C-/D+/D- grades. They stopped using these grades in 2016.

Proposed Change:

Simply eliminate the C-/D+/D- lines from this chart, and include a note beneath the chart. "Beginning with the fall 2023 semester, UCF eliminated the C-/D+/D- grades."

Grades	Grading Points Per Semester Hour of Credit
A	4.00
A-	3.75
B+	3.25
B	3.00
B-	2.75
C+	2.25
C	2.00
C-	1.75
D+	1.25
D	1.00
D-	.75
F	0.00
NC - No Credit	*

Recommendations for Carnegie Elective Classification for Community Engagement

Recommendation 1: Invest in a centralized database to track and align community engagement efforts
and assess how they help or potentially hinder success of the university.

Recommendation 2: Invest in annual data collection processes across the university to track community
engagement efforts and outcomes.

Recommendation 3: Identify resources to conduct community impact assessments of community engaged scholarship activities.

Recommendation 4: Launch a fund through UCF Foundation for community engaged faculty and staff to
leverage, by application, to compensate community participants, particularly from marginalized communities, for their participation in research, teaching, and creative activities with the university.

Recommendation 5: Utilize existing structures, offices, and initiatives to provide capacity building and
training for faculty, staff, and students who are implementing or want to implement community engaged scholarship and practice.

Recommendation 6: Encourage colleges and subunits to explore tenure and promotion standards that
recognize and reward faculty for community engaged scholarship.

Recommendation 7: Create a position (e.g. Associate Vice Provost) and/or an office (e.g. Social Impact and Community Engagement) to coordinate, monitor, assess, and strengthen community engaged scholarship and practices throughout the university.