WHITE PAPER

Identifying Some Sources of Bias in Course and Instructor Evaluations (CIEs)

Updated:

March 6, 2021

Prepared by the Faculty Affairs Committee

DRAFT REPORT[1]

[1] This informational report is the work of the members of the Faculty Affairs Committee and is not the official policy of Rollins College.

ACKNOWLEDGEMENTS

The Faculty Affairs Committee wishes to extend its appreciation to Professor Benjamin Hudson for his work preparing an earlier draft of this document. Also, the Committee wishes to thank Dr. Nancy Chick for supplying important references used in the preparation of this report.

FACULTY AFFAIRS MEMBERSHIP

Missy Barnes (2020-2022)
Dr. David Caban (2019-2021)
Dr. Ashley Cannaday (2019-2021)
Dr. Leigh DeLorenzi (2020-2022)
Dr. John Grau (2018-2020)
Dr. Benjamin Hudson (2018-2020)
Dr. Margaret McLaren (2020-2022)
Dr. Leslie Poole (2019-2021)
Dr. Samuel Sanabria (2019-2021)
Dr. Rachelle Yankelevitz (2019-2021)
Dr. Donald Davison, chair (2019-2021)
Dean Jennifer Cavenaugh, Ex Officio

PREFACE

The Rollins College Faculty Affairs Committee (FAC) was requested by several faculty members and academic administrators to re-examine the efficacy of the current online course instructor evaluation (CIE) method. The course instructor evaluation tool serves as one important part of the evaluation of teaching effectiveness at Rollins College. Like any subjective rating process, the CIE is limited because it can reflect users' racial and gender biases. This White Paper is an initial examination of evaluating teaching effectiveness surveyed in the national literature as well as at Rollins College. Accordingly, the FAC recommends ongoing analysis of teaching effectiveness and possible sources of bias.

To that end, this White Paper examines the phenomena of racial, gender and sexual orientation bias in CIEs. Nonetheless the FAC does not recommend abolishing CIEs. Instead we ask evaluators to be aware of possible bias and encourage more effective use of the CIE. The intention behind this White Paper is to provide an educational resource to faculty and administrators about the limitations of course evaluations in evaluating faculty for tenure and promotion. While course evaluations can provide valuable feedback to a faculty member on how to improve her or his courses and can also reveal areas of strengths and weaknesses in teaching, best practices indicate that course evaluations should be only one measure of a variety of measures to evaluate teaching. There is a prolific literature examining the reliability and validity of student evaluations of teaching (SET) in higher education. Generally, the literature reports the robust conclusion that online course evaluations are vulnerable to biases correlated with gender, race, and sexual orientation of the instructor. In addition, the literature generally finds that many course evaluations are poor measures of student learning. Instead, the instruments tend to capture student satisfaction with the course, their perception of learning rather than actual learning, and their grade expectations. Course evaluations can reflect students' (sometimes implicit) biases and as such may often be impoverished sources of information about minority and female faculty in administrative review of teaching effectiveness.

This White Paper provides an overview of the national literature regarding gender, race, and sexual orientation-related biases in course evaluation. We also identify some of the unique characteristics of Rollins College which separate us from other institutions in these studies.

Next, we report general descriptive results regarding the outcomes from the CIEs at Rollins as they compare to the trends found in the literature. Finally, the goal of the FAC is to prepare recommendations that will be discussed with the faculty during the spring, 2021. Excellence in teaching is the *sine qua non* of Rollins College. As a faculty we are eager to inform ourselves of our teaching effectiveness and student learning. We hope to increase awareness of the strengths and limitations of course evaluations thus encouraging a forum for discussion and development.

Course instructor evaluations (CIEs) play a significant role in career trajectories, in both personnel and awards decisions for faculty at many institutions, including Rollins. A chorus of recent inquiries into the efficacy of course evaluations across various institutions suggests that they may provide limited information about teaching effectiveness generally, and they frequently can reflect the unconscious biases of students. The limitations of course evaluations are magnified in the context of evaluating minority faculty. This white paper examines gender, racial, and sexual biases, although other sources of bias exist. The literature affirms the importance of using a holistic approach for evaluating teaching that recognizes the limitations of course evaluations and includes other measures of evaluating teaching.

GENERAL LIMITATIONS OF TEACHING EVALUATIONS

Since the 1990s, when course evaluations began to take on significant importance in hiring, retention, and promotion decisions at American universities, scholars have sounded the alarm on their efficacy.[2] In a recent 2017 review of the literature, and which includes some strong suggestions for rethinking course evaluations, Henry Hornstein notes several problems with standardizing the evaluation of teaching. These problems include: (1) considerable disagreement about what qualities mark "teaching effectiveness" and the problem of measurement generally; (2) a reminder that CIEs are objectively suspect because they measure students' subjective perceptions of a course and instructor rather than the actual course and instructor herself; (3) the problem of limited response rates; and (4) that student satisfaction does not necessarily correlate with learning. Hornstein surveys the ways in which course evaluations do not offer a solid

[2] See, for example, J.V. Adams, "Student Evaluations: The Ratings Game." *Inquiry* 1 (1997): 10-16.

ground on which instruction can be measured objectively. In response, he suggests that "the persistent practice of using student evaluations as summative measures to determine decisions for retention, promotion, and pay for faculty members is improper and depending on circumstances could be argued to be illegal."[3]

Many studies conclude that course evaluations are flawed measures of teaching effectiveness.[4] Boring, et. al., find that student evaluations are more strongly related to the instructor's gender and to students' grade expectations than objective indicates of learning. "On the whole, high SET (student evaluations of teaching) seem to be a reward students give instructors who make them anticipate getting a good grade. . . ."[5] Boring and her colleagues also find gender disparities in student teaching evaluations. Overall, male instructors receive higher scores than female instructors. However, they also find gender concordance—male students give male instructors higher evaluation scores than they give female instructors, and vice versa. Therefore, gender effects may be heightened depending on the composition of the instructor's class. For instance, a female instructor with a largely male student class might expect to receive statistically significant lower evaluations regardless of how much learning occurred in the course. Indeed, Deslauriers and colleagues found little relationship between perceived learning and objective learning in introductory physics classes.[6] The authors found that students who are engaged in active learning—while more difficult than passive learning—demonstrate objectively greater knowledge on end of the year exams. Consistent with this objective, Rollins College encourages active learning by students even though it is more challenging. Despite the advantages of active learning, however, some students may perceive themselves to learn more under passive learning approaches. This could lead to a disconnect between the effectiveness of a course measured by student learning and the perceptions held by students revealed in the course evaluation.

---

[3] Hornstein, Henry, "Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance." *Cogent Education* 4 (2017): 1-8, 2.
[4] Boring, Anne, Kellie Ottoboni, and Philip Start, "Student evaluations of teaching (mostly) do not measure teaching effectiveness," *ScienceOpen Research,* January 7, 2016.
[5] Ibid, p. 1.
[6] Deslauriers, Louis, Logan McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin, "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom," *PNAS Latest Articles*, August 13, 2019.

Finally, Esarey and Valdes use computational simulation that assumes course evaluations are valid, reliable, and unbiased. They find that even under these ideal assumptions course evaluations cannot reliably identify good teaching. Instead, they recommend that using course evaluations in combination with multiple measures of teaching effectiveness can produce better results.[7]

The FAC would like to add that course evaluations for courses that involve controversial, emotionally triggering, or political content might confuse indicators of student learning with student perceptions of a class. This might be especially true for faculty from underrepresented groups who teach about topics related to their identity, for example, African American faculty who teach about racism and white privilege.

GENDER BIAS IN TEACHING EVALUATIONS

A robust scholarship over the last thirty years indicates that student evaluations unfairly critique the teaching effectiveness of female instructors due not to "gendered behavior" on behalf of the instructors but to "actual bias on the part of the students."[8] In a 2015 study from MacNell, Driscoll, and Hunt, the authors emphasize that student gender biases reflect a broader trend of "the pervasive devaluation of women, relative to men, that occurs in professional settings in the United States" (293). The authors show that gender bias in course evaluations is a significant source of inequality facing female faculty and "systematically disadvantages women in academia" (301).

Ben Schmidt, professor of history at Northwestern University, has compiled data from over 14 million Ratemyprofessor.com reviews in interactive graphs on his professional website that reveal the unconscious bias of student evaluations. According to Claire Cain Miller, Schmidt's data reveals "that people tend to think more highly of men than women in professional settings,

---

[7] Esarey, Justin and Natalie Valdes, "Unbiased, reliable, and valid student evaluations can still be unfair," *Assessment and Evaluation in Higher Education*, February 20, 2020.
[8] MacNell, Lillian, Adam Driscoll, and Andrea Hunt, "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40 (2015): 291-303, 301. Subsequent references appear parenthetically within the text.

praise men for the same things they criticize women for, and are more likely to focus on a woman's appearance or personality and on a man's skills and intelligence."[9] Schmidt's visualizations of his data, available on his website show significant discrepancies along gender lines in student evaluations of teaching: male instructors are more likely to be rated "smart," "genius," or "funny," while female professors are more frequently labeled "strict" or "bossy." Professor Schmidt's frequency analysis of RateMyProfessor.com is limited in that Ratemyprofessor.com tends to attract a nonrepresentative sample of course evaluators; however, its strength is that the site is possibly the largest publicly-available database of course evaluations.

More recently, scholars Kristina Mitchell and Jonathan Martin demonstrate the differences in language students use to evaluate male and female faculty. They show that a male instructor "administering an identical course as a female instructor receives higher ordinal scores in teaching evaluations, even when questions are not instructor-specific."[10] Mitchell and Martin demonstrate that student evaluations of female faculty often demean their professional accomplishments, critique their attire and personality, and generally document "that students have less professional respect for their female professors" (652). These data encourage Mitchell and Martin to argue against course evaluations in administrative or promotional decisions altogether because "the use of evaluations in employment decisions is discriminatory against women" (648).

RACIAL AND ETHNIC BIAS IN TEACHING EVALUATIONS

Although course evaluations have existed in higher education for nearly a century, it is no surprise that education researchers have historically "overlooked the classroom experiences of

---

[9] Miller, Claire Cain, "Is the Professor Bossy or Brilliant? Much Depends on Gender.," *New York Times*, 6 Feb. 2015.

[10] Mitchell, Kristina M. and Jonathan Martin, "Gender Bias in Student Evaluations." *PS: Political Science & Politics* 51, 3 (July 2018):, 648-652, 648. Subsequent references appear parenthetically within the text.

teachers and professors of color."[11] Over the last several decades, this lacuna has begun to be addressed as education researchers have investigated the challenges facing professors of color in regards to the validity of course evaluations and the instrument's tendency to reflect prejudices. Thirty years ago, textile and clothing scholar Usha Chowdhary conducted two different sections of the same course in different garb—one in traditional Indian clothing and the other in Western clothing; she discovered that the course evaluations from the section in which she wore traditional Indian clothing were more negative.[12] Ten years later, Heidi Nast surveyed "student resistances to multicultural teaching and faculty diversity [and] the risks that derive from problematic institutional deployment of student evaluations as a means of judging multicultural curricular and faculty success."[13] Nast surveys several incidents when course evaluations were used to harass faculty of color and/or LGBTQ faculty and "to register anger and disapproval at having to negotiate topics and issues in a scholarly way which conflict with heretofore learned social values and assumptions" (104). A contemporaneous study by Katherine Hendrix similarly determines that "race influences student perceptions of professor credibility" (740) and that "the competence of Black professors was more likely to be questioned" (758). This review only scratches the surface of a robust scholarship from the end of the twentieth century; Chowdhary, Nast, and Hendrix help us understand how course evaluations for classes taught by faculty of color frequently reflect larger social biases and are this must be weighed when using course evaluations as a measure of success in the classroom.[14]

While Chowdary, Nast, and Hendrix relied on anecdotal data from restricted sample sizes, more recently scholars have broadened the scope of their investigations. In a robust review of evaluations from students at 25 liberal arts colleges on the website *Ratemyprofessor.com*,

[11] Hendrix, Katherine Grace, "Student Perceptions of the Influence of Race on Professor Credibility." *Journal of Black Studies* 28, 6 (1998): 738-763, 739. Subsequent references appear parenthetically within the text.
[12] Chowdhary, Usha, "Instructor's Attire as a Biasing Factor in Students' Ratings of an Instructor." *Clothing & Textiles Research Journal* 6 (1988): 17-22.
[13] Nast, Heidi J, "'Sex', 'Race' and Multiculturalism: Critical Consumption and the Politics of Course Evaluations." *Journal of Geography in Higher Education* 23, 1 (03, 1999): 102-115, 103. Subsequent references appear parenthetically within the text.
[14] A more recent study confirms their findings: Arnold K Ho, Lotte Thomsen, and Jim Sidanius,. "Perceived Academic Competence and Overall Job Evaluations: Students' Evaluations of African American and European American Professors." *Journal of Applied Social Psychology* 39.2 (2009): 389-406.

Landon Reid determined that "racial minority faculty, particularly Black faculty, were evaluated more negatively than White faculty in terms of Overall Quality, Helpfulness, and Clarity."[15] Reid cautions that "both race and gender have an interactive effect on course evaluations that should be considered in the tenure and promotion cases of racial minority faculty" (145). Importantly, Reid points out that students "are unlikely to assert that a racial minority faculty member is a bad instructor because of their race" and that "instead, prejudicial biases are more likely to be expressed as principled, and therefore socially defensible, evaluations of an instructor's teaching" (146). Reid noted particularly that at institutions like Rollins, which "demand excellent, not merely good, teaching for promotion and tenure" the problem of racial minority faculty's evaluative disadvantage may be "compounded" (148).

Similarly, Bettye Smith and Billy Hawkins contribute to the discussion with a large-scale quantitative, empirical study which determined that "race does matter in how students evaluate both faculty and the value of the courses faculty teach […] and therefore matters when examining faculty effectiveness."[16] Smith and Hawkins's study demonstrates that Black faculty's "mean scores were the lowest" among Black, White, and a third racial category of Other (159). Smith and Hawkins find that this phenomenon was "especially troublesome because these ratings have the power to affect merit increases and careers" (159). Other studies have addressed this evaluative disadvantage shouldered by minority faculty, with similar findings that Hispanic and Asian American faculty similarly receive lower ratings than White faculty.[17]

SEXUAL ORIENTATION BIAS IN TEACHING EVALUATIONS

---

[15] Reid, Landon, "The Role of Perceived Race and Gender in the Evaluation of College Teaching on RateMyProfessors.com." *Journal of Diversity in Higher Education* 3, 3 (2010): 137-152, 145. Subsequent references appear parenthetically within the text.
[16] Smith, Bettye P. and Billy Hawins, "Examining Student Evaluations of Black College Faculty: Does Race Matter?" *The Journal of Negro Education* 80, 2 (2011): 149-162, 160. Subsequent references appear parenthetically within the text.
[17] Anderson, K.J. and Smith, G. "Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender." *Hispanic Journal of Behavioral Sciences*, 2 (2005):184-201; and G. Smith, G and Anderson, K.J,. "Students' Ratings of Professors: The Teaching Style Contingency for Latino/a Professors." *Journal of Latinos and Education* 4 (2005): 115-136.

There is a growing literature investigating whether students' evaluations of professors are influenced by their perception of the faculty member's sexual orientation. Generally, conclusions about students' racial and gender biases extend to biases about sexual orientation of instructors. For instance, Melanie Moore and Richard Trahan find that women who teach courses on gender often experience resistance and skepticism because students perceive them as advancing their personal political agenda.[18] By extension, Russ, Simonds, and Hunt (2002) examine whether instructor sexual orientation influences students' perceptions of teacher credibility, character, and students' personal assessment of how much they are learning.[19] Their results suggest that perceptions of credibility, character, and student learning are strongly influenced by the sexual orientation of the instructor. In comparing student ratings of a guest instructor who indicated he was either gay or straight, "Students perceived the gay instructor to be significantly less credible in terms of competence and character" compared to their evaluations of the straight instructor (316). Similarly, analyzing qualitative information such as written comments revealed that the gay instructor vignette received four-times more negative comments by students compared to the straight instructor. Russ and Simonds also reveal a connection between students' perception of how much they learn, the credibility of the instructor, and the sexual orientation of the instructor. First, they find that students perceive themselves to learn more from teachers who are seen as credible. Second, "students perceive they learn almost twice as much from a heterosexual teacher compared to a gay teacher (319)." In summary, students rate a gay instructor as less credible and therefore perceive themselves as learning less than from a heterosexual instructor.

In addition to perceived learning perceptions, Kristin Anderson and Melinda Kanner report that "Lesbian and gay professors were rated as having a political agenda, compared to heterosexual

[18] Moore, Melanie and Richard Trahan, "Biased and political: Student perceptions of females teaching about gender." *College Student Journal*, 31, 4, (1997).
[19] Russ, Travis L. Cheri J. Simonds, and Stephen K. Hunt, "Coming Out in the Classroom . . . An Occupational Hazard?: The Influence of Sexual Orientation on Teacher Credibility and Perceived Student Learning," *Communication Education,* 51, 3, (2002).

professors with the same syllabus (1538).[20] These results suggest that students' course evaluation criteria differ when evaluating courses taught by lesbian or gay professors versus heterosexual professors. This expanding body of literature shows that there are biases regarding the sexual orientation of instructors.

COURSE INSTRUCTOR EVALUATIONS AT ROLLINS COLLEGE

The current course and instructor evaluation instrument (CIE) was adopted in 2007. The CIEs provide several unique and important sources of information for the instructor of the course and the evaluation committees. The CIE provides longitudinal information regarding a faculty member's development as a teacher. In this way, the instrument offers information about the patterns and trajectories of faculty teaching; the CIEs also provide narrative feedback from student comments. The qualitative information from student comments can be combined with the numeric information available from the inventory of evaluation areas receiving scaled scores. Both qualitative and quantitative information can be useful to faculty members to reflect upon and improve their teaching and for evaluation committees to identify patterns and areas of concern. As this White Paper discusses, course evaluations may reflect bias in both the narrative comments and numerical scores. We should recognize that the CIEs at Rollins are subject to some of the limitations associated with all teaching evaluation instruments used at institutions across the United States. Because of this Rollins should carefully consider the role of course evaluations in tenure and promotion decisions and ensure that we use a holistic approach for evaluating teaching which includes course evaluations, syllabi, assignments, exercises, simulations, classroom observation, etc. The evaluator should combine the qualitative student comments and the quantitative scores to gain a narrative and numeric picture of the students' perceptions of the course.

BIAS AT ROLLINS

The Office of Institutional Analytics examined whether there is evidence of bias in the quantitative component of the Course and Instructor Evaluation (CIE) instrument used at Rollins.

---

[20] Anderson, K. J., & Kanner, M., Inventing a gay agenda: Students' perceptions of lesbian and gay professors. *Journal of Applied Social Psychology, 41*(6), 1538–1564, (2011). https://doi.org/10.1111/j.1559-1816.2011.00757.x

The study was conducted using 1,837 course sections taught by full-time CLA faculty from fall, 2016 through fall, 2019. This produced a pool of more than 32,000 separate course evaluations used in the statistical analysis. International faculty and faculty who did not specify their race or ethnicity in the College survey are excluded from the analysis. The results indicate very small differences in the quantitative scores between male and female faculty as well as between white non-Hispanic faculty and faculty from minority groups.

Two different analyses were conducted. The first test compared the difference in mean raw scores for each indicator in the CIE between faculty groups. The differences in mean raw scores range from 0.02 to 0.10 of one raw score point (significant; $p<0.05$). The second analysis examined the difference in the percentage of course evaluations that receive either a Poor (score = 1) or Fair (score=2) on items in the inventory. In other words, this analysis explores the possibility that certain groups of faculty receive a larger number of extremely poor evaluations compared to their white male colleagues. The results show that female faculty and faculty from minority groups receive 0.40% to 1.50% more evaluations with low scores (significant; $p<.05$). (Refer to Appendix for complete results).

This analysis had a very large sample size (N). Large-N studies such as this can sometimes produce an illusory statistical significance, such that even though the statistical tests are significant, they may only appear that way due to the large sample size. This, even when groups do not truly differ, they may "significantly" differ when N is very large. Therefore, the FAC requested two additional analyses. First, the Office of Institutional Analytics calculated effect size (Cohen's d). The measure of effect size compares any two groups to see how much they differ from each other. Cohen's d is a statistic used to measure the standardized difference between two means. A 'rule of thumb' is when d is less than 0.2 it indicates small differences between the sample means. When d approaches 0.5 there is evidence of a moderate effect and when d approaches 0.8 the effect is considered large. In our data set, the majority of the comparisons have a d less than 0.2 although a few items range between 0.2 and 0.4. (See Appendix 3). Thus, the effect sizes (the differences between the groups) were small to moderate. This analysis of quantitative differences (statistical significance and effect size) in quantitative dimension of the CIE does not address the potential psychological or evaluative impact of the numbers on the perceptions and actions of evaluators and instructors themselves, nor does it address bias in students' comments.

Finally, Appendix 4 reports the results for whether there was a difference in the average size of class enrollments by the faculty groups. If faculty from under-represented groups or female faculty members regularly teach classes that are larger (smaller) compared to white (male) faculty then there could be a class size effect influencing the results. The results indicate that class sizes are comparable across all groups in the study and this test provided no evidence of a class-size effect.

Summary Comparison of Quantitative CIE Scores For Faculty Groups		
		Range (min – max differences)
Minority Faculty compared with White faculty		
	Range of mean differences in raw scores (minority means < white means)	0.02 – 0.10
	Range of difference in percent of evaluations either Poor (1) or Fair (2) (minority percent > white percent)	0.53% - 1.47%
Female compared with Male Faculty		
	Range of mean differences in raw scores (female < male)	0.02 – 0.09
	Range of percent of evaluations either Poor (1) or Fair (2) (female percent > male percent)	0.39% - 1.45%
29,733 < N <32,307		

The faculty of Rollins College strive to be excellent teachers. Faculty value the information they receive from their course evaluations each semester as they reflect on and fine-tune their classes. The Faculty Affairs Committee offers several recommendations designed to heighten awareness of the subtle ways bias influences course evaluations as well as ways to best use the information contained in the CIEs. The FAC hopes these suggestions will increase awareness of the potential forms of bias and contribute to a discussion of how to effectively evaluate teaching in liberal arts colleges.

1. The Office of Institutional Analytics should conduct the Race and Gender Bias Study every four years and report the results to the Faculty Affairs Committee. We recommend that the next study also include an analysis of student comments. This enables an analysis of both quantitative and qualitative information contained in the evaluations. Regular reporting of this information allows faculty and administrators to monitor the institution's progress regarding resisting bias in teaching evaluations and aids in effectively using the information contained in the CIEs.

2. The FAC recommends that the text box for faculty comments on the CIE is made a permanent feature on Course Instructor Evaluations.

3. The FAC recommends that the name of the instrument be changed from Course Instructor Evaluation to "Student Perceptions of the Course and Instruction."

4. The FAC encourages faculty to view the online tutorial available for using the CIE). The instructional tutorial is very thorough and provides useful contextual information for properly interpreting course evaluations, possible biases in raw scores and comments, and interpretation of the comparison percentiles.

5. CIEs can provide useful longitudinal information by identifying trends and patterns in faculty instruction. The strategy for interpreting CIEs is combining the quantitative measures (raw scores) with the qualitative information available in students' comments. The FAC affirms that a holistic approach to evaluation is preferrable in which CIEs are combined with other sources of information about teaching quality and development.

6. The FAC recommends that evaluators avoid relying on the percentiles except when they reveal a consistent pattern below the 10th percentile. The overall distribution of teaching scores at Rollins is very high. Therefore, small changes in raw scores can produce large changes in the corresponding percentile score.

Appendix

Results for Negative Bias against Female Faculty and Faculty from Unrepresented Groups

Negative Rating Bias Against Female Faculty in Student Course Evaluations

Chi-square test for Equal proportions

Null Hypothesis H0 = Both female and male faculty are equally likely to receive negative rating (1=Poor and 2=Fair) from student

i.e. H0 = the proportions of negative rating received by male and female faculty = 0.5

Alternate Hypothesis H1 = Male and female faculty are not equally likely to receive negative rating from a student

For each of questions below, where **p-value < 0.05**, reject the null hypothesis and infer that the proportion of negative ratings received by male and female faculty are not equal

Conclusion: This study shows that full-time Female Faculty at Rollins College consistently receive more negative rating in student course evaluations compared to their male counterpart

		for Female Faculty					for Male Faculty								
#	Survey Question	Responses of 1 Poor and 2 Fair	% of Responses of 1 Poor and 2 Fair	Responses of 3 Good, 4 Very Good and 5 Excellent	% of Responses of 3 Good, 4 Very Good and 5 Excellent	Total # of Responses	Responses of 1 Poor and 2 Fair	% of Responses of 1 Poor and 2 Fair	Responses of 3 Good, 4 Very Good and 5 Excellent	% of Responses of 3 Good, 4 Very Good and 5 Excellent	Total # of Responses	Difference in % of 1 Poor and 2 Fair responses (Male - Female)	Chi-Square Statistic Value	Prob or p-value	N
11.2	Overall Professor - Overall, how would you rate this professor?	1,140	6.8%	15,745	93.2%	16,885	812	5.3%	14,514	94.7%	15,326	-1.45%	29.81	4.8E-08	32,211
7. Please rate your professor on the following characteristics-															
7.1	Respectful - Treats students with courtesy and respect	467	2.7%	16,521	97.3%	16,988	336	2.2%	15,036	97.8%	15,372	-0.56%	10.58	1.1E-03	32,360
7.2	Prepared - Organized & prepared when teaching students	900	5.3%	16,063	94.7%	16,963	557	3.6%	14,787	96.4%	15,344	-1.68%	52.52	4.3E-13	32,307
7.3	Enthusiastic - Genuinely excited about teaching & interacting with students	366	2.2%	16,589	97.8%	16,955	321	2.1%	15,021	97.9%	15,342	-0.07%	0.17	6.8E-01	32,297
7.4	Effective - Able to explain complex material & accomplish course goals	936	5.5%	16,012	94.5%	16,948	749	4.9%	14,588	95.1%	15,337	-0.64%	6.65	9.9E-03	32,285
7.5	Interesting - Draws your interest & keeps your attention	1,151	6.8%	15,802	93.2%	16,953	975	6.4%	14,368	93.6%	15,343	-0.43%	2.47	1.2E-01	32,296
7.6	Knowledgeable - Comprehensive & current knowledge in her/his field	323	1.9%	16,626	98.1%	16,949	194	1.3%	15,138	98.7%	15,332	-0.64%	20.95	4.7E-06	32,281
7.7	Egalitarian - Treats students equally - does not play favorites	716	4.2%	16,214	95.8%	16,930	517	3.4%	14,801	96.6%	15,318	-0.85%	15.95	6.5E-05	32,248
7.8	Tolerant - Open to student attitudes & opinions that are not her/his own	730	4.3%	16,081	95.7%	16,811	508	3.3%	14,736	96.7%	15,244	-1.01%	21.96	2.8E-06	32,055
7.9	Supportive - Encourages students to do their best & supports their efforts	575	3.4%	16,334	96.6%	16,909	461	3.0%	14,862	97.0%	15,323	-0.39%	3.97	4.6E-02	32,232
7.10	Available - Easy to approach & available for meetings outside of class	712	4.3%	15,768	95.7%	16,480	514	3.4%	14,394	96.6%	14,908	-0.87%	15.88	6.8E-05	31,388

Lower Average Score Bias Against Female Faculty in Student Course Evaluations

Two sample t-test for Equal Average Scores

Null Hypothesis H0 = The avg. score given by students to male and female faculty are equal (or statistically indifferent). Avg. score for each faculty is calculated for each of the below questions asked in student course evaluation by considering the following scores: 1 for Poor, 2 for Fair, 3 for Good, 4 for Very Good and 5 for Excellent.

Alternate Hypothesis H1 = Average scores given to male and female faculty by the students in course evaluation is not equal.

For each of questions below, where **Probt < 0.05**, reject the null hypothesis and infer that the average score received by the male and female faculties in that question is not the same.

Conclusion: This study shows that full-time Female Faculty at Rollins College consistently receive a lower average score in student course evaluations compared to their male counterpart

#	Survey Question	Average Score of Female Faculty (mu1)		Average Score of Male Faculty (mu2)	Difference between Avg. Score of Male - Female Faculty	Method	Variances	tValue	DF	Probt	Method	Variances	tValue	DF	Probt
11.2	Overall Professor - Overall, how would you rate this professor?	4.37	<	4.46	0.09	Pooled	Equal	-21.60	32,209	<.0001	Satterthwaite	Unequal	-21.69	32,203	<.0001
7. Please rate your professor on the following characteristics-															
7.1	Respectful - Treats students with courtesy and respect	4.66	<	4.70	0.04	Pooled	Equal	-17.32	32,358	<.0001	Satterthwaite	Unequal	-17.51	31,932	<.0001
7.2	Prepared - Organized & prepared when teaching students	4.50	<	4.59	0.09	Pooled	Equal	-24.26	32,305	<.0001	Satterthwaite	Unequal	-24.54	31,732	<.0001
7.3	Enthusiastic - Genuinely excited about teaching & interacting with students	4.69	<	4.71	0.02	Pooled	Equal	-8.04	32,295	<.0001	Satterthwaite	Unequal	-8.00	30,973	<.0001
7.4	Effective - Able to explain complex material & accomplish course goals	4.48	<	4.53	0.05	Pooled	Equal	-16.76	32,283	<.0001	Satterthwaite	Unequal	-16.78	32,103	<.0001
7.5	Interesting - Draws your interest & keeps your attention	4.42	<	4.47	0.05	Pooled	Equal	-18.49	32,294	<.0001	Satterthwaite	Unequal	-18.42	31,332	<.0001
7.6	Knowledgeable - Comprehensive & current knowledge in her/his field	4.72	<	4.79	0.07	Pooled	Equal	-35.41	32,279	<.0001	Satterthwaite	Unequal	-35.67	32,219	<.0001
7.7	Egalitarian - Treats students equally - does not play favorites	4.60	<	4.65	0.05	Pooled	Equal	-20.72	32,246	<.0001	Satterthwaite	Unequal	-20.85	32,224	<.0001
7.8	Tolerant - Open to student attitudes & opinions that are not her/his own	4.59	<	4.66	0.07	Pooled	Equal	-23.06	32,053	<.0001	Satterthwaite	Unequal	-23.25	31,857	<.0001
7.9	Supportive - Encourages students to do their best & supports their efforts	4.65	<	4.67	0.02	Pooled	Equal	-11.46	32,230	<.0001	Satterthwaite	Unequal	-11.48	32,033	<.0001
7.10	Available - Easy to approach & available for meetings outside of class	4.59	<	4.64	0.05	Pooled	Equal	-14.74	31,386	<.0001	Satterthwaite	Unequal	-14.83	31,366	<.0001

** The above study was conducted by the Office of Provost with results collected from student course evaluations in CLA courses from most recent 7 Spring and Fall terms (Fall 2016 through Fall 2019) for 1,837 sections taught by our current 200 full-time CLA faculty. The analysis was carried out on the 11 questions asked to students in course evaluations that rate faculty on their teaching and behavior in the classroom. The four groups used for this analysis are full-time female faculty, full-time male faculty, full-time faculties from White Non-Hispanic race and faculties from Under-represented Minority (URM) races. URM group includes faculty from Asian, African American race and, Hispanic ethnicity. International faculty and faculty who have not specified their Race or Ethnicity to the college survey have been excluded from the study. All race, ethnicity and gender categories are self-identified by the individuals.

Negative Rating Bias Against Under-represented Faculty in Student Course Evaluations

Chi-square test for Equal proportions

Null Hypothesis H0 = Both Under-represented faculty (URM) and White Non-Hispanic faculty are equally likely to receive negative rating (1=Poor and 2=Fair) from students

i.e. H0 = the proportions of negative rating received by URM and White faculty = 0.5

Alternate Hypothesis H1 = URM and White faculty are not equally likely to receive negative rating from a student

For each of questions below, where **p-value < 0.05**, reject the null hypothesis and infer that the proportion of negative ratings received by URM and White faculty are not equal

Conclusion: This study shows that full-time Faculties from Under-represented Races at Rollins College consistently receive a more negative rating in student course evaluations compared to other White Non-Hispanic Faculties

		for Under-represented (URM) Faculty					for White Non-Hispanic Faculty								
#	Survey Question	Responses of 1 Poor and 2 Fair	% of Responses of 1 Poor and 2 Fair	Responses of 3 Good, 4 Very Good and 5 Excellent	% of Responses of 3 Good, 4 Very Good and 5 Excellent	Total # of Responses	Responses of 1 Poor and 2 Fair	% of Responses of 1 Poor and 2 Fair	Responses of 3 Good, 4 Very Good and 5 Excellent	% of Responses of 3 Good, 4 Very Good and 5 Excellent	Total # of Responses	Difference in % of 1 Poor and 2 Fair responses (White - URM)	Chi-Square Statistic Value	Prob or p-value	N
11.2	Overall Professor - Overall, how would you rate this professor?	346	7.2%	4,449	92.8%	4,795	1,450	5.6%	24,264	94.4%	25,714	-1.58%	18.14	2.1E-05	30,509
7. Please rate your professor on the following characteristics-															
7.1	Respectful - Treats students with courtesy and respect	139	2.9%	4,684	97.1%	4,823	593	2.3%	25,237	97.7%	25,830	-0.59%	5.99	1.4E-02	30,653
7.2	Prepared - Organized & prepared when teaching students	236	4.9%	4,583	95.1%	4,819	1,084	4.2%	24,702	95.8%	25,786	-0.69%	4.73	3.0E-02	30,605
7.3	Enthusiastic - Genuinely excited about teaching & interacting with students	142	2.9%	4,679	97.1%	4,821	475	1.8%	25,300	98.2%	25,775	-1.10%	24.99	5.8E-07	30,596
7.4	Effective - Able to explain complex material & accomplish course goals	304	6.3%	4,512	93.7%	4,816	1,234	4.8%	24,534	95.2%	25,768	-1.52%	19.72	9.0E-06	30,584
7.5	Interesting - Draws your interest & keeps your attention	350	7.3%	4,471	92.7%	4,821	1,616	6.3%	24,161	93.7%	25,777	-0.99%	6.63	1.0E-02	30,598
7.6	Knowledgeable - Comprehensive & current knowledge in her/his field	99	2.1%	4,717	97.9%	4,816	377	1.5%	25,389	98.5%	25,766	-0.59%	9.30	2.3E-03	30,582
7.7	Egalitarian - Treats students equally - does not play favorites	218	4.5%	4,586	95.5%	4,804	932	3.6%	24,813	96.4%	25,745	-0.92%	9.41	2.2E-03	30,549
7.8	Tolerant - Open to student attitudes & opinions that are not her/his own	212	4.4%	4,563	95.6%	4,775	923	3.6%	24,668	96.4%	25,591	-0.83%	7.76	5.3E-03	30,366
7.9	Supportive - Encourages students to do their best & supports their efforts	210	4.4%	4,598	95.6%	4,808	745	2.9%	24,977	97.1%	25,722	-1.47%	28.94	7.5E-08	30,530
7.10	Available - Easy to approach & available for meetings outside of class	198	4.3%	4,452	95.7%	4,650	936	3.7%	24,147	96.3%	25,083	-0.53%	2.96	8.5E-02	29,733

Lower Average ScoreBias Against Under-represented Faculty in Student Course Evaluations

Two sample t-test for Equal Average Scores

Null Hypothesis H0 = The avg. score given by students to URM and White Non-Hispanic faculty are equal (or statistically indifferent). Avg. score for each faculty is calculated for each of the below questions asked in student course evaluation by considering the following scores: 1 for Poor, 2 for Fair, 3 for Good, 4 for Very Good and 5 for Excellent.

Alternate Hypothesis H1 = Average scores given to URM and White faculty by the students in course evaluation is not equal.

For each of questions below, where **Probt < 0.05**, reject the null hypothesis and infer that the average score received by the URM and White faculties in that question is not the same.

Conclusion: This study shows that full-time Faculty from Under-represented Races at Rollins College consistently receive a lower average score in student course evaluations compared to other White Non-Hispanic Faculty

#	Survey Question	Average Score of URM Faculty (mu1)		Average Score of White Non-hispanic Faculty (mu2)	Difference between Avg. Score of White - URM Faculty	Method	Variances	tValue	DF	Probt	Method	Variances	tValue	DF	Probt
11.2	Overall Professor - Overall, how would you rate this professor?	4.37	<	4.44	0.07	Pooled	Equal	-8.72	30,507	<.0001	Satterthwaite	Unequal	-7.67	6,083	<.0001
7. Please rate your professor on the following characteristics-															
7.1	Respectful - Treats students with courtesy and respect	4.66	<	4.69	0.03	Pooled	Equal	-9.63	30,651	<.0001	Satterthwaite	Unequal	-8.83	6,296	<.0001
7.2	Prepared - Organized & prepared when teaching students	4.54	<	4.56	0.02	Pooled	Equal	-4.9	30,603	<.0001	Satterthwaite	Unequal	-4.53	6,332	<.0001
7.3	Enthusiastic - Genuinely excited about teaching & interacting with students	4.66	<	4.72	0.06	Pooled	Equal	-11.75	30,594	<.0001	Satterthwaite	Unequal	-8.74	5,578	<.0001
7.4	Effective - Able to explain complex material & accomplish course goals	4.44	<	4.54	0.10	Pooled	Equal	-14.5	30,582	<.0001	Satterthwaite	Unequal	-12.11	5,916	<.0001
7.5	Interesting - Draws your interest & keeps your attention	4.41	<	4.48	0.07	Pooled	Equal	-6.52	30,596	<.0001	Satterthwaite	Unequal	-5.48	5,947	<.0001
7.6	Knowledgeable - Comprehensive & current knowledge in her/his field	4.73	<	4.77	0.03	Pooled	Equal	-7.81	30,580	<.0001	Satterthwaite	Unequal	-6.46	5,882	<.0001
7.7	Egalitarian - Treats students equally - does not play favorites	4.62	<	4.63	0.02	Pooled	Equal	-10.52	30,547	<.0001	Satterthwaite	Unequal	-9.6	6,246	<.0001
7.8	Tolerant - Open to student attitudes & opinions that are not her/his own	4.60	<	4.64	0.04	Pooled	Equal	-10.81	30,364	<.0001	Satterthwaite	Unequal	-9.59	6,091	<.0001
7.9	Supportive - Encourages students to do their best & supports their efforts	4.61	<	4.67	0.06	Pooled	Equal	-16.73	30,528	<.0001	Satterthwaite	Unequal	-13.21	5,726	<.0001
7.10	Available - Easy to approach & available for meetings outside of class	4.59	<	4.63	0.04	Pooled	Equal	-11.34	29,731	<.0001	Satterthwaite	Unequal	-9.39	5,679	<.0001

** The above study was conducted by the Office of Provost with results collected from student course evaluations in CLA courses from most recent 7 Spring and Fall terms (Fall 2016 through Fall 2019) for 1,837 sections taught by our current 200 full-time CLA faculty. The analysis was carried out on the 11 questions asked to students in course evaluations that rate faculty on their teaching and behavior in the classroom. The four groups used for this analysis are full-time female faculty, full-time male faculty, full-time faculties from White Non-Hispanic race and faculties from Under-represented Minority (URM) races. URM group includes faculty from Asian, African American race and, Hispanic ethnicity. International faculty and faculty who have not specified their Race or Ethnicity to the college survey have been excluded from the study. All race, ethnicity and gender categories are self-identified by the individuals.

Appendix 3

Effect Size (Cohen's D)

			Gender	Female			Male			Grand Total			
Question Num	Question Title	Question Order	Question Text	Total Responses Female Faculty	Mean Score	Std. Dev.	Total Responses Male Faculty	Mean Score	Std. Dev.	Total Responses	Mean Score	Std. Dev. Population	Effect Size
11	Please rate your professor on the following characteristics	2	Overall Professor - Overall, how would you rate this professor?	16,885	4.3747	0.3969	15,326	4.4607	0.3553	32,211	4.4130	0.3804	0.2260
7	Please rate your professor on the following characteristics	1	Respectful - Treats students with courtesy and respect	16,988	4.6589	0.2411	15,372	4.7028	0.2019	32,360	4.6784	0.2250	0.1955
		2	Prepared - Organized & prepared when teaching students	16,963	4.4968	0.4032	15,344	4.5868	0.3110	32,307	4.5368	0.3669	0.2453
		3	Enthusiastic - Genuinely excited about teaching & interacting with students	16,955	4.6885	0.2234	15,342	4.7070	0.2557	32,297	4.6967	0.2378	0.0778
		4	Effective - Able to explain complex material & accomplish course goals	16,948	4.4829	0.3441	15,337	4.5279	0.3218	32,285	4.5029	0.3343	0.1347
		5	Interesting - Draws your interest & keeps your attention	16,953	4.4227	0.3360	15,343	4.4736	0.3745	32,296	4.4453	0.3536	0.1439
		6	Knowledgeable - Comprehensive & current knowledge in her/his field	16,949	4.7212	0.1846	15,332	4.7900	0.1556	32,281	4.7518	0.1752	0.3925
		7	Egalitarian - Treats students equally - does not play favorites	16,930	4.5967	0.2476	15,318	4.6527	0.2148	32,248	4.6216	0.2346	0.2386
		8	Tolerant - Open to student attitudes & opinions that are not her/his own	16,811	4.5929	0.2733	15,244	4.6614	0.2250	32,055	4.6234	0.2546	0.2691
		9	Supportive - Encourages students to do their best & supports their efforts	16,909	4.6490	0.2286	15,323	4.6691	0.2258	32,232	4.6580	0.2269	0.0884
		10	Available - Easy to approach & available for meetings outside of class	16,480	4.5891	0.2707	14,908	4.6376	0.2298	31,388	4.6107	0.2539	0.1911

			Race	URM			White (non-Hispanic)			Grand Total			
Question Num	Question Title	Question Order	Question Text	Total Responses	Mean Score	Std. Dev.	Total Responses	Mean Score	Std. Dev.	Total Responses	Mean Score	Std. Dev. Population	Effect Size
11	Please rate your professor on the following characteristics	2	Overall Professor - Overall, how would you rate this professor?	4,795	4.3714	0.4183	25,714	4.4428	0.3638	30,509	4.4314	0.3727	0.1916
7	Please rate your professor on the following characteristics	1	Respectful - Treats students with courtesy and respect	4,823	4.6640	0.2300	25,830	4.6905	0.2169	30,653	4.6863	0.2186	0.1210
		2	Prepared - Organized & prepared when teaching students	4,819	4.5410	0.3678	25,786	4.5569	0.3512	30,605	4.5544	0.3528	0.0452
		3	Enthusiastic - Genuinely excited about teaching & interacting with students	4,821	4.6611	0.3180	25,775	4.7209	0.2001	30,596	4.7113	0.2231	0.2679
		4	Effective - Able to explain complex material & accomplish course goals	4,816	4.4412	0.4029	25,768	4.5391	0.3018	30,584	4.5234	0.3208	0.3053
		5	Interesting - Draws your interest & keeps your attention	4,821	4.4083	0.4279	25,777	4.4796	0.3225	30,598	4.4682	0.3411	0.2090
		6	Knowledgeable - Comprehensive & current knowledge in her/his field	4,816	4.7313	0.2017	25,766	4.7660	0.1569	30,582	4.7604	0.1647	0.2106
		7	Egalitarian - Treats students equally - does not play favorites	4,804	4.6156	0.2410	25,745	4.6325	0.2279	30,549	4.6298	0.2294	0.0736
		8	Tolerant - Open to student attitudes & opinions that are not her/his own	4,775	4.5997	0.2688	25,591	4.6412	0.2363	30,366	4.6346	0.2415	0.1720
		9	Supportive - Encourages students to do their best & supports their efforts	4,808	4.6138	0.2897	25,722	4.6747	0.2074	30,530	4.6650	0.2227	0.2735
		10	Available - Easy to approach & available for meetings outside of class	4,650	4.5885	0.2925	25,083	4.6262	0.2378	29,733	4.6201	0.2468	0.1525

Appendix 4

Class Size Effects

		Number of classes	Avg Class Size	StdDev	Q1	Median	Q3
By race	URM	436	15.07	5.43	11	15	19
	White (non-Hispanic)	2236	16	5.75	12	16	21
By gender	Female	1513	15.3	5.41	11	15	20
	Male	1305	16.54	5.97	12	17	21