SIRS Analysis Results

Executive Summary

While recent scholarship on student feedback about teaching (or student evaluations of teaching) that suggests that there may be gendered or racial differences in student feedback (see Kreitzer & Sweet-Cushman, 2021) has garnered national attention, not all studies have found evidence of such differences. Rutgers is unique in that, by many measures, the student body is among the country's most diverse of large public institutions, which may impact student feedback on teaching. This report includes several analyses requested by the Rutgers University Senate, including differences in how students rate an instructor depending on the instructor's gender or race, correlations between Student Instructional Ratings Survey [SIRS] questions, and factor analysis. Additional research on Rutgers faculty and administrator perceptions and use of SIRS is also included.

For the primary analysis, Spring 2022 Student Instructional Ratings data was used as a sample. This does not include any units using the Banner registration system (some of the RBHS units) or Nursing clinical surveys but includes the majority of SIRS with 133,254 student responses and a total of 5,151 instructors. The following are the main findings from each analysis and related literature.

Initial Analysis

Distributions demonstrated all ten items on the standard survey were highly skewed and tended to be more positive than negative.

- The mean of SIRS9 (teaching effectiveness) was 4.29, and the mean of SIRS10 (course quality) was 4.15 [both on a 5-point Likert scale from Poor (1-point) to Excellent (5-points)].
- Skewness and kurtosis statistics were found to deviate significantly from normality.

Differences by Gender of Instructor

Much of the current studies look at institutions' student survey results overall, which is where we began. The sample includes slightly more male instructors (52%) than female instructors (45.5%).

- University-wide, females were rated higher than males on both SIRS9 (Females M = 4.33; Males M = 4.25), the instructor effectiveness question, and SIRS10, the overall course quality question (Females M = 4.18; Males M = 4.11). While these are small differences, at less than a tenth of a point each, both are statistically significant ($p < 0.001$).
- These differences (0.08 for teaching effectiveness and 0.07 for the overall course quality question), are not consistent with other studies, some of which have found that men score higher than women on similar questions. Boring and colleagues (2016) found that men were rated 0.08 higher than women university wide at a French university. A study at McGill University (2018) in Canada observed a 0.02 difference in favor of men. Another study from a diverse Southwest university found no statistically significant difference between women and men (Miles & House, 2015).

Comparisons of means across the university and schools were conducted to align with other research in this area. However, tenure, promotion, and advancement decisions take place at the department level, and therefore, it is

important to consider if there are differences at that level. Given the size and diversity of Rutgers University, a comparison across schools or the entire university is of limited use in that it compares courses and instructional methods that are wildly different, which is partly why a unit or university-wide mean is not included in SIRS results.

- The size and direction of differences in student ratings of male and female instructors depend on the school and department.
- In some departments, female instructors have a higher mean while in other departments, male instructors have a higher mean. Ratings for females and males were not statistically different in other departments. While there is some evidence that male instructors may be rated more highly in STEM departments due to the culture of the discipline (Cashin, 1990; Linse, 2017), this was not seen in our sample.

Differences by Race of Instructor

Instructor race/ethnicity was grouped into 'White' and 'Nonwhite', similar to Chávez and Mitchell's (2020) study to provide a larger pool of instructors. The university cannot require ethnicity or race, and since it is a voluntary field, instructors can choose to include or not.

- Due to the voluntary self-identification of race/ethnicity, 46.6% of instructors at the time of analysis did not indicate race/ethnicity.
- Due to this large number of unknowns, additional analysis is not meaningful.

Correlations

Since the data is not normal, Kendall's Tau was used for the bivariate correlation. Statistical significance was calculated, but statistical significance is expected with such large sample sizes.

- Strong positive correlations, all statistically significant at the .01 level, were found between all questions except for SIRS8 (the question asking about their prior interest in the course).
- The correlation coefficient between SIRS9 (teaching effectiveness) and SIRS10 (course quality) was 0.73.
- The full correlation matrix can be found in the slides.

Factor Analysis

The item about students' prior interest (SIRS8) has weak positive correlations but is also an aspect not connected (or negatively connected possibly) to students' perception of teaching effectiveness. Therefore SIRS8 was not included in the factor analysis.

- The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) indicated factor analysis should produce reliable and distinct factors (KMO = .947).
- All KMO values for individual items were greater than .90, which is well above the acceptable limit of .5 (Field, 2013), and Bartlett's test was significant ($p < .001$).
- The principle axis factor analysis resulted in one factor with an eigenvalue of 6.92 and high reliability with a Cronbach's alpha of .96. Factor loadings can be found in the slides.
- The result of only one factor is in line with the literature, which has found student surveys measure students' individual and general learning experiences in their course (Spooren, Brockx, & Dimitri, 2013). Some studies did uncover multiple factors, but these were surveys with 35 or more items (Jackson et al., 1999).
- A regression analysis was not conducted due to the high correlations between the various questions and the uncovering of only one factor.

Perceptions of SIRS: OTEAR Research

Instructors' Perceptions of SIRS

Questions pertaining to the value of SIRS were included within the Spring 2023 Instructor Survey administered by OTEAR and received 1,322 responses (24% response rate).

- Instructors indicated they find SIRS results valuable for formative uses (improving their course, M = 3.77 out of 5 being Strongly Agree). In contrast, not as many see the SIRS results as helpful for summative processes (promotion, tenure, and advancement, M = 3.17).
- This was also supported in the comments, as most instructors who found student feedback useful told us they used it to improve their courses. This ranged from gauging the effectiveness of an assignment, a video, a resource, or an activity to improving the presentation of content or the alignment between content and assessment materials.
- Those instructors who do not find student feedback valuable provided several reasons, including a belief that students provided biased responses, the belief that the feedback was never helpful, and concern over low response rates.

Administrators' Perceptions of SIRS

Individuals who review SIRS for their programs, departments, or schools were asked to complete a survey and participate in a focus group during the 23-24 academic year to understand how SIRS results were used as part of the evaluation of teaching. Eight-seven respondents included individuals from all four chancellor-led units.

- 73% of respondents indicated they strongly agree or agree that student feedback provides useful evidence of the quality of instruction.
    - Example from a comment: "When many negative open-ended comments corroborate low quantitative scores, they both flag and provide guidance about how to intervene with struggling instructors. In one case, we asked for a course redesign based on claims about disorganization, and supported the instructor in developing a redesigned course."
- Respondents recognize the limitations inherent in student feedback and suggested various strategies they use when interpreting student feedback:
    - When there are a few extremely negative or critical responses that are not in line with feedback from a majority of students, these responses should be discounted.
    - When the response rate for a survey is very low and a few negative or critical responses are received, the entire survey may be discounted and evaluators should focus on other evidence.
    - When the response rate is strong and a substantial number of students express negative or critical responses, an evaluator should seek additional evidence to see whether it corroborates student concerns, including by giving instructors opportunities to respond and to make adjustments to instruction if possible. Even when substantial student dissatisfaction is expressed, additional evidence may lead an evaluator to decide that the teaching methods and approach meet the standards of the department.

Limitations & Considerations

Some additional items to consider include:

- Questions ranged from 1.9% missing responses (for SIRS10) to 7.9% of the responses missing SIRS5. These were excluded from the analyses which utilized responses to all questions.
- Schools have the ability to not use the Standard SIRS form as long as they retain SIRS9 (Overall Teaching Effectiveness) and SIRS10 (Overall Course Quality) questions.
    - For example, Pharmacy recently redesigned their surveys to include tailored questions for each of the various types of courses (basic science and general courses, iPASS, Pharmacotherapy, and clinical elective). Their thoughtful process of developing questions and piloting the forms took approximately two years and they are currently evaluating them using various measures.

Implications

The analyses indicate that the current SIRS questions may provide an overall indication of students' perception of their experience in a course. Faculty and administration value SIRS while acknowledging its limitations. Where do we go next? Here are some current directions of some of our peers:

- Adding a statement to the beginning of the survey to mitigate unconscious bias (Genetin et al., 2021; Peterson et al., 2019).
- Focusing on how to communicate the purpose of SETs and what is actionable feedback with training and videos (Signorini et al., 2020)
- Customizing end-of-course student surveys to improve the quality of student feedback (some Rutgers schools have done this as well as detailed in the last section)
- Strengthening holistic evaluation of teaching through peer review, portfolios, and other evidence (Dennin et al., 2017)

Support Available from OTEAR

OTEAR supports the implementation of SIRS and the holistic evaluation of teaching including:

- Resources to support instructors in communicating about SIRS & providing feedback - https://go.rutgers.edu/CommunicateSIRS
- Workshops on interpreting SIRS, peer review observations, and teaching portfolios for faculty and administrators - https://otear.rutgers.edu/workshops/
- Consultations with schools regarding customizing their forms, piloting, and evaluation - https://otear.rutgers.edu/sirs/sirs-forms/

References

Bangert, A. W. (2008). The development and validation of the student evaluation of online teaching effectiveness. Computers in the Schools, 25(1-2), 25-47.

Boring, A., Ottoboni, K., & Stark, P.B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen research*. doi: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Cashin, W.E. (1990). Students do rate different academic fields differently. New Directions for Teaching and Learning, 43, 113-121.

Chávez, K., & Mitchell, K. M. (2020). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science & Politics*, *53*(2), 270-274.

Dennin, M., Schultz, Z. D., Feig, A., Finkelstein, N., Greenhoot, A. F., Hildreth, M., Leibovich, A. K., Martin, J. D., Moldwin, M. B., O'Dowd, D. K., Posey, L. A., Smith, T. L., & Miller, E. R. (2017). Aligning Practice to Policies: Changing the Culture to Recognize and Reward Teaching at Research Universities. CBE life sciences education, 16(4), es5. https://doi.org/10.1187/cbe.17-02-0032

Dziuban, C., & Moskal, P. (2011). A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments. The Internet and Higher Education, 14(4), 236–241. https://doi.org/10.1016/j.iheduc.2011.05.003

Field, A. (2013). Discovering statistics using IBM SPSS statistics. *Reference & Research Book News, 28(2)*. Ringgold, Inc.

Genetin, B., Chen, J., Kogan, V., & Kalish, A. (2021). Mitigating implicit bias in student evaluations: A randomized intervention. Applied Economic Perspectives and Policy, 44(1), 110-128. doi: https://doi.org/10.1002/aepp.13217

Jackson, D. L., Teal, C. R., Raines, S. J., Nansel, T. R., Force, R. C., & Burdsal, C. A. (1999). The dimensions of students' perceptions of teaching effectiveness. *Educational and Psychological Measurement*, *59*(4), 580-596.

Kreitzer, R.J., & Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. Journal of Academic Ethics, 20:73–84. https://doi.org/10.1007/s10805-021-09400-w

Linse, A.R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. Studies in Educational Evaluation, 54, 94-106.

McGill University. (2018). Report on Gender Equity in Mercury Course Evaluations at McGill. McGill University. https://www.mcgill.ca/mercury/channels/news/report-gender-equity-mercury-course-evaluations-mcgill-289553

Miles, P., & House, D. (2015). The Tail Wagging the Dog; An Overdue Examination of Student Teaching Evaluations. International Journal of Higher Education, 4(2), 116–126.

Peterson, D. A., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. PLoS One, 14(5), e0216241. doi: https://doi.org/10.1371/journal.Pone.0216241

Signorini, A., Abuan, M., Panakkal, G., & Dorantes, S. (2020). Students helping students provide valuable feedback on course evaluations. To Improve the Academy: A Journal of Educational Development, 39(2), 75-105. doi: https://doi.org/10.3998/tia.17063888.0039.204

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. https://doi.org/10.3102/0034654313496870