ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #36369

WIP: Identity-Based Bias in Undergraduate Peer Assessment

Miss Madison Jeffrey, University of Michigan

Madison Jeffrey is a graduate candidate in the University of Michigan's Masters in Higher Education program. With a focus on Management and Organizations, she's interested in ways in which the system of higher education can adapt to become more accessible and equitable to students. She's a research assistant at the University of Michigan's College of Engineering, where she works with a team of researchers responsible for Tandem, a software that monitors team performance to link students and instructors.

Dr. Robin Fowler, University of Michigan

Robin Fowler is a lecturer in the Program in Technical Communication at the University of Michigan. She enjoys serving as a "communication coach" to students throughout the curriculum, and she's especially excited to work with first year and senior students, as well as engineering project teams, as they navigate the more open-ended communication decisions involved in describing the products of open-ended design scenarios.

Mark Mills, UM, Center for Academic Innovation

Mark Mills is a Data Scientist with the Center for Academic Innovation at the University of Michigan. He is responsible for leading analysis across the Center in support of its mission to leverage data for shaping innovation in higher education. Mark received his PhD from the University of Nebraska in Cognitive and Quantitative Psychology, where he studied models for classifying cognitive state from eye movements.

WIP: Identity-Based Bias in Undergraduate Peer Assessment

Madison Jeffrey
Center for the Study of Higher & Postsecondary Education,
University of Michigan

Robin Fowler
Center for Academic Innovation & College of Engineering,
University of Michigan

Mark Mills
Center for Academic Innovation,
University of Michigan

Introduction

Peer assessments are commonly found across classrooms that have a focus on small-group learning and, occasionally, are used to influence the grade a student receives in the course. The practice of using peer assessment is common because of its use in assessing an individual's success and performance [1-3]. When peer assessment is used as a tool to determine the academic efforts of a student, it is important to understand the implicit processes that impact such decisions [4]. This paper is concerned with identifying trends in peer assessments that are related to the race and gender of the rater and ratee involved.

Others have studied this question with mixed results concerning the potential role bias may play in impacting ratings and how bias can appear as trends in statistical differences in the ratings students are given [5-10]. An explanation for statistical differences between identities, whether it be gender, race, or economic background, is that biases of the rater are impacting ratings. Studies of gender bias in peer assessment have mixed results. Some studies have found that no significant difference exists at all for gender [5-6], even in qualitative analyses [7]. One study finds that in an undergraduate economics course women rated women 50% higher than men rated women, but, in another course within the same study, there were no statistical differences [8]. There is less literature on racial biases that shows significant differences [9], but studies have found non-significant trends in lower ratings for historically underrepresented students [10].

Other individual differences have been found to impact peer ratings such as personality [11] and '*perceiver variance*' [12], in which 32% of the variance in peer ratings was due to consensus on the ratee, 20% was due to rater tendency to score high or low on certain characteristics, and other variation was residual. The residual variance could be due to relationship-specific variance, such as raters having "distinctive biases" towards ratees [12]. The residual variance could present itself as gender or racial biases, which don't account for the skill of an individual but rather the *perception* a rater has of an individual from perceived identities.

Research Question & Data

The goals of this study are to gain a deeper understanding of how potential biases are influencing student ratings. In this preliminary analysis, we first consider whether *differences* in race/ethnicity or gender between the rater and ratee have a statistically significant effect on received ratings. Understanding whether these effects are present in the sample will then allow us to see the relationship of the identities and frame the next step of our study.

This study uses data gathered in semesters from Fall 2019 to Fall 2020 from a web-based tool, Tandem, used in some courses at our institution to monitor team performance [13]. The courses represented in this sample of the data are twelve undergraduate courses (8 Engineering, 2

Business, 1 Architecture, and 1 Digital Design) enrolling 1,789 total students at a large Midwestern research institution. Completion of the peer assessment survey was incentivized via course credit in most courses, yielding a high response rate. This research will help to inform the software program in its peer assessment construction and will add to the growing literature on peer evaluation. This is especially important within STEM fields with an over-representation of certain genders, races, and socioeconomic identities [14] [15].

Students report their gender and race/ethnicity at the start of the course through Tandem. The racial distribution of this sample is 55.5% white, 31.4% Asian, 2.83% Black, 4.77% Latino, 2.49% Middle Eastern, 0.27% Native American, and 0.23% Pacific Islander. Students who Self-Described their racial identity made up 0.39% of the sample and 0.21% were Undisclosed. In this sample, 56.9% identified as male, 42.5% identified as female, and 0.56% as non-binary.

Methods

In this preliminary analysis, we conducted a hierarchical linear regression to predict the outcome of peer-rated values using racial and gender identities. Peers rated one another on the following characteristics: brought valuable skills to the project (*et_valuable*), actively taught others (*et_teacher*), showed up reliably (*et_reliable*), created exceptional output (*et_quality*), listened to others well (*et_listener*), offered many ideas (*et_ideas*), provided ideas that were used in the project (*et_enacted*), did a fair share of the work (*et_effort*), and seemed to belong on the team (*et_belonging*). All characteristics were measured using a 1 to 9 Likert scale, with 9 being positive. The dependent variable, *value*, is the numeric value of the characteristics assigned by the rater to the ratee. To look at racial and gender differences between rater and ratee, we used independent variables of *gender_diff* and *race_diff*. A score of '1' indicates that the gender or race of the rater and ratee differed and a score of '0' indicates the gender or race matched.

The preliminary regression model accounted for a variety of covariates. These included rater's high school GPA (*hsgpa*), rater's estimated gross family income (*est_inc*) , and rater's mean cumulative GPA (*cum_gpa*), as recorded in the university database. An index, *peer*, represented the average of the rater's scores of each member of the group in areas such as the peers' ideas, effort, quality, and reliability. Controlling for these covariates allowed us to examine the increase in the significance of the model fit when we introduced the two independent variables of interest. We also included *rater* (a student's ID as a rater), *ratee* (a student's ID as a ratee), and the aforementioned *characteristics*. These three variables used random effects to account for individual differences of each student as rater and ratee. Other variables were treated as fixed.

Results and Discussion

The model controls for GPA (high school *and* cumulative current), economic status, and the overall peer-group scores assigned by the rater, With the model's conditional variance of .616, the model explains 61.6% of the variance in rating values.

There was a statistically significant difference in values given to ratees when the rater and ratee had different genders ($p=.006$, $b=-.025$) and indicates that when raters and ratees had different genders, their values given to ratees would decrease by .25 points on the 9-point Likert scale compared with raters and ratees of the same gender. For instances in which raters and ratees had a different race or ethnicity ($p=.002$, $b=.024$), values given to ratees would increase by .24 points

on the 9-point Likert scale compared with raters and ratees of the same race or ethnicity. Student ratees and raters being of different race/ethnicities had a significant effect on values given to the ratees. In short, being a different gender from one's peer can negatively impact ratings given to peers in this sample of engineering and business students, both of which are male-dominated. Being a different race or ethnicity from one's peer can positively impact ratings given to peers.

The model above does not differentiate between races being assessed and the relationships between them. It is possible that the effects of bias stem from implicit, stereotyped expectations of others and that the effects of that bias act in predictable ways depending on the races involved, rather than only occurring if students don't share a racial identity. In this sample, 86.9% of participants were white or Asian, making all other racial identities a minority of the sample at 13.1%. Seeing a positive effect of racial or ethnic differences on ratings given to peers, which was seen in the above model, could potentially be an outcome of students being exposed to new experiences and perspectives, or of students having internalized biases against people who share their own identity. Without parsing the individual relationships, we cannot be certain that there aren't negative or positive effects of race/ethnicity dependent upon the identities involved.

Future Directions

Simplifying racial or ethnic identities does not allow us to understand the nuances of racial biases that may impact student perception of behavior. The approach of simplifying peer identities to different or same doesn't consider the individuals' specific identity or if that identity is historically underrepresented. Understanding how biases impact the students' perception of their own task efficacy is integral in peer assessment as we aim to make STEM fields such as engineering and, more generally, higher education a more equitable experience for all students regardless of background.

We hope to next analyze directional differences within dyads. A dyadic analysis would allow us to consider the specific race and gender of the ratee *and* how the identities of the rater interact with them to impact ratings. This approach also provides details on what identities are receiving significantly different ratings from their peers. This approach will also show if those significantly different ratings are being received from a group of peers with specific identities. We plan to consider the characteristics that students are being rated on and if certain identities are scored higher or lower in a characteristic compared to others. Understanding trends across identities and accounting for group variances [11] will provide a better understanding of what impacts ratings beyond individual variance, and we can gain this understanding in the next steps of the study.

Highlighting the potential conflicts present in peer assessment would serve to advance the important equity efforts universities are undertaking across the nation. As more research is conducted showing the ways in which assessment ratings can be swayed by characteristics and perceptions of students, instructors must approach the practice with a complete understanding of what those ratings reflect. Establishing best practices for peer assessment in how it is conducted, determining the subject matter, and how it is reviewed is integral to the growth of small learning practices and its positive impacts on the student experience.

References

[1] M. Donia, T. O'Neill, & S. Brutus. (2018). The longitudinal effects of peer feedback in the development and transfer of student teamwork skills. In *Learning and Individual Differences, 61*, 87-98.

[2] C. Brooks & J. Ammons. (2003). Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments. In *Journal of Education for Business, 78*(5), 268-272.

[3] J. Beatty, R. Haas, & D. Sciglimpaglia. (1996). Using peer evaluations to assess individual performance in group class projects. In *Journal of Marketing Education, 18*, 17-27.

[4] Sridharan, B., Tai, J., Boud, & D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? In *Higher Education, 77*, 853-870.

[5] D. Kaufman, R. Felder, & H. Fuller. (2000). Accounting for individual effort in cooperative learning teams. In *Journal of Engineering Education*, *89*(2), 133-140.

[6] R. Tucker. (2014). Sex does not matter: gender bias and gender differences in peer assessments of contributions to group work. In *Assessment & Evaluation in Higher Education, 39*(3), 293-309.

[7] P. Birch, J. Batten, & J. Batey. (2015). The influence of student gender on the assessment of undergraduate student work. In *Assessment & Evaluation in Higher Education, 41*(7), 1065-1080.

[8] M. Espey. (2021). Gender and peer evaluations. In *The Journal of Economic Education, 53*(1), 1-10.

[9] J. Ghorpade & J. Lackritz. (2001). Peer evaluation in the classroom: a check for sex and race/ethnicity effects. In *Journal of Education for Business, 76*(5), 274-281.

[10] D. Kaufman, R. Felder, & H. Fuller. (1999). Peer ratings in cooperative learning teams. In *Proceedings of the 1999 annual ASEE Meeting*, 1-11.

[11] G. May & L. Gueldenzoph. (2000). The effect of social style on peer evaluation ratings in project teams. In *International Journal of Business Communication*, *43*(1), 4-20.

[12] C. Martin & K. Locke. (2022). What do peer evaluations represent?: a study of rater consensus and target personality. In *Frontiers in Education, 7*, 1-7.

[13] The Regents Of The University Of Michigan, "Tandem," tandem.ai.umich.edu. https://tandem.ai.umich.edu/welcome (accessed: June 23, 2022).

[14] ASEE. (2019). Current status of the U.S. engineering and computing workforce, 2019.

[15] J. Roy. (2019). Engineering by the Numbers. *American Society for Engineering & Education*, 13-52.