International Journal of Engineering Education Vol. 40, No. 3, pp. 543–556, 2024
Printed in Great Britain

0949-149X/91 $3.00+0.00

Evaluation of Bias in Peer Assessment in Higher Education*

JACKLIN H. STONEWALL and MICHAEL C. DORNEICH**
Industrial and Manufacturing Systems Engineering, Iowa State University, 3028 Black Engineering Building, Ames, IA 50011-2164, USA. Email: dorneich@iastate.edu

JANE RONGERUDE
Community and Regional Planning, Iowa State University, Ames, IA, USA.

Small group, active learning strategies have increased student achievement, attendance, engagement, and overall learning outcomes. Peer assessment is an important aspect of team pedagogy. It increases individual accountability, improves team functioning, and contributes to a sense of belonging among teammates. However, perceptions of unfairness undermine many positive outcomes of team learning and peer assessment. Students worry that their peers are not impartial raters; indeed, research has shown that they are not. However, a broad investigation of bias is needed. Specifically, there is little peer assessment data that cuts across classes, departments, and academic fields, as this information tends to be challenging to collect into a single dataset. This project aimed to explore bias in peer assessment from multiple perspectives through a literature review, surveys of students and instructors, and an analysis of peer assessment ratings given and received. The first study broadly asked instructors about the occurrence of bias in their classrooms and actions to mitigate it. The second study asked students about their peer assessment experiences and perceived biases. The third study analyzed over 20,000 peer assessment ratings to investigate bias. Both instructors and students detailed bias in their classrooms and with assessments. Evidence of bias was shown in the peer assessment scores, which student achievement cannot fully explain. The results demonstrated that the experiences of women and students of color in these classrooms differ from those of their peers regarding assessment. By understanding where and how bias occurs in peer assessment, training can be designed to target problem areas and improve the fairness of assessment directly. This could ensure that the positive outcomes associated with learning teams are shared among all students.

Keywords: team-based learning; student assessment; bias; active learning; teams

1. Introduction

The employment of small group, active learning strategies (such as cooperative learning or Team-Based Learning) in classroom environments has been shown to increase student achievement, attendance, and engagement and to result in better overall learning outcomes [1–3]. Engineering classrooms using active learning have demonstrated enhanced cognitive acquisition of material over conventional lecturing approaches [4, 5], with even more significant benefits for students from underrepresented groups [6]. Because of these outcomes, team-based pedagogies and cooperative learning practices have been identified within institutions of higher learning as a strategy to improve the classroom engagement and retention of underrepresented students. Research shows that learning in teams positively affects objective outcomes (such as exam scores) for minority students [7, 8].

In engineering, the process of peer assessment is suggested as a method of teaching and a technique for reinforcing core professional skills (e.g., communication, peer work review, and team skills). As a result, it is not surprising that a growing number of engineering professional associations now recommend that engineering education programs integrate active and small-group learning strategies into their curriculum. These organizations include the European Society for Engineering Education (SEFI), the Active Learning in Engineering Education (ALE) network, and accreditation organizations such as the Accreditation Board for Engineering and Technology (ABET) [5]. Negative experiences with teaming and peer assessment can potentially leave all engineering students professionally unprepared [9].

Peer assessment is often included within a small team pedagogy. In peer assessment, students evaluate the quality of their colleagues' classroom engagement, contribution, and effort, and it often counts toward the students' final course grades [10, 11]. Peer assessments are usually anonymous and supported by detailed instructions and a grading rubric [12, 13]. Graded assessments reduce "social loafing" (failing to participate), increase individual accountability [14], and contribute to a sense of belonging among teammates [15].

Many of the positive outcomes of team learning and peer assessment can be undermined by perceptions of unfairness [14, 16]). Studies report that students consider the peer assessment process fair *in principle* [16]. However, students worry that, *in practice*, their peers are not impartial raters, and

** Corresponding author.

* Accepted 12 February 2024.

indeed, research has shown that they are not [16–18]. Biases that create unfairness in peer assessment have been observed in the classroom [19] and in assessment scores themselves [20]. These observed biases have been attributed to gender, race, ethnicity, socioeconomic status, and peer group affiliation.

Unfortunately, some engineering students from underrepresented groups encounter a chilly classroom climate. Biases have been shown to manifest in many ways, including in peer assessments [21], which can contribute to a sense of not being welcome and can deter diverse students from continuing their engineering education [22, 23]. By understanding where biases in peer assessment occur in both qualitative and quantitative feedback, educators can be better positioned to address them and enable all students to reap the benefits of active learning and peer assessment.

Researchers have examined the occurrence of bias in peer assessment in individual classes (e.g., [21, 24]) or departments (e.g., [25]). In one academic year, 38% of professors surveyed perceived an act of bias in their classes [19]. However, a broad investigation of bias is missing from the literature. Specifically, little of the existing peer assessment data cuts across classes, departments, and academic fields as this information tends to be difficult to collect into a single dataset. There are further gaps for studies that examine the issue of peer assessment bias from multiple perspectives (e.g., student perceptions, instructor perceptions, and peer assessment scores themselves).

This paper reports on findings from three studies investigating bias in peer assessment from three perspectives: instructor, student, and peer assessment rating data. All studies were conducted in the context of one large US Midwestern university and were approved by the Institutional Review Board (IRB). Study 1 asked instructors about the occurrence of bias in their classrooms and in peer assessments and any actions they had taken to mitigate bias. Study 2 asked students about their peer assessment experiences and perceived biases. Study 3 analyzed a large body of peer assessment data collected over five years across multiple departments and disciplines to determine if and when bias occurred.

2. Literature Review

2.1 Benefits of Peer Assessment

A growing body of research demonstrates the value of peer assessment for the learning process [26]. Instructors report using peer assessment as a tool for increasing students' motivation [27–29], promoting the learning process [12, 30] and increasing students' engagement [31]. While increased accountability receives the most attention, peer assessments have been demonstrated to empower learners and increase their classroom engagement [32], increase interactions among students and between students and instructors [33], and contribute to a sense of belonging within teams [15]. Assessments can foster autonomy and maturity and help students improve social and professional skills [34]. These skills transfer to the workplace, where implementing peer review and peer feedback allows employees to understand better their strengths and weaknesses [35].

Recently, the Accreditation Board for Engineering and Technology (ABET) revised the student outcomes required for accreditation. Outcome 5 now reads: *"an ability to function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives."* [36]. With this more detailed definition of the team environment, ABET suggests evidence can be obtained through peer assessment: *"Use of web-based peer evaluations such as CATME.org or TEAMMATES. The peer evaluations include specific questions about collaboration and inclusiveness."* ([36], pp. 5). Student Outcome 5 also specifically prescribes that the team environment be inclusive. As teaming and peer assessment are not always inclusive processes for all students, this work on improving the fairness of peer assessments can be directly applied to the satisfaction of Student Outcome 5.

Despite these benefits, students and instructors have expressed concerns about the fairness of peer assessment [37–40]. The primary threat to peer assessment fairness is bias, which may be explicit or implicit [41]. Biases that may impact peer assessment have been observed in the classroom environment (e.g., [19]) and in assessment scores themselves (e.g., [20]).

2.2 Fairness in Peer Assessments

Peer evaluations have become the focus of a growing body of research, including examinations of student perceptions [17] and peer assessment implementation strategies [42]. Both students and instructors have expressed concerns about the fairness of teams and their associated peer assessments, primarily due to concerns with bias [37–40]. It is already understood that biased behaviors are commonly present in higher education classrooms [19]. Research has shown that women and students of color also have different experiences in terms of assessment [21]. The challenge in creating peer assessments that maximize fairness and minimize biases is two-sided: (1) the difficulty of detecting

various forms of bias in peer assessment and (2) the difficulty of making students aware of their biases or potential for bias.

2.3 Threats to Peer Assessment Fairness

Biased actions result from an individual's biased attitudes, which can be implicit or explicit. These attitudes create an individual's subjective organizational structure for perceiving their environment [41]. Although biased attitudes are deeply engrained, they tend to be inconsistently expressed depending on the social context. For example, explicit attitudes are often moderated or "censored" in social situations where they are perceived as unwelcome [43]. Classroom bias tends to target individuals' perceived sexual orientation, race, gender, and ethnicity [44].

Implicit bias refers to an attitude that is "... activated by the mere presence (actual or symbolic) of the attitude object and commonly functions without a person's full awareness or control" ([41], p. 62). In the classroom, implicit bias manifests in varying ways. For example, one study of bias in university classrooms divided occurrences into three categories of microaggressions: microassaults (exclusion), microinsults (subtle verbal snubs largely unknown to the perpetrator), and microinvalidations (negating the experiences of marginalized groups) [19, 38]. Role assignment within classroom teams can be biased, with men assigned more technical roles and females more secretarial roles [45]. Bias has also been identified in the marks that instructors give to students. Female students have been shown to receive lower class participation scores than male students despite no evidence for this disparity in other aspects of the course (e.g., exam and homework scores) [39].

Explicit bias is a consciously held belief about a person or group. Where implicit attitudes are challenging to self-recognize, control, and measure, explicit attitudes are overt and more readily measured by traditional assessment measures. A study on the experiences of South Asian students in predominantly white classrooms found that all of the 40 students interviewed had experienced explicit bias. [38]. Instructors report implicit and explicit bias incidents occurring at similar frequencies in the classroom [19]. The incidences of explicit bias most reported by professors are the explicit use of stereotypes, offensive jokes, and racial and ethnic slurs [19].

2.4 Bias in Peer Assessment

Many students were concerned that when completing peer assessments, it would be "difficult to avoid personal bias" ([46], p. 85). While the classroom biases discussed have focused on gender, race, and ethnicity, peer assessment bias has also been observed due to social style, socioeconomic status, native language, sexual orientation, and social entanglement [47].

Results have been mixed in studies that focus on the overall ratings *received* by male and female students. Multiple researchers have found that female students receive lower ratings and fewer positive qualitative comments than their male peers [48–50]. However, other studies have shown that males receive lower ratings than their female peers [20, 24, 25, 51]. The mixed results extend to analyses of the ratings *given* by male and female students. In some cases, women have been found to give higher evaluation scores than men [51]. Conversely, another study found that men give higher evaluation scores and women give lower marks [49]. A study of Team-Based Learning (TBL) general education classes showed that while gender bias in the assessment scores was not observed, women did more work in team activities, suggesting that their extra work was going unrewarded [21].

Student attitudes toward peer assessment are also inconsistent. A study of undergraduates found no gender-based differences in satisfaction with the peer assessment process [52]. However, more recent research has shown that male students report more positive attitudes about peer assessment than female students [33, 53]. One of the few consistent findings for gender effects relates to ratings given by students to themselves. Female students consistently underrate themselves [54–56], while male students consistently overrate themselves [54, 55].

Studies of racial bias in peer assessment have returned mixed results. Older studies have found that individuals tend to be rated higher by members of their own race than those of other races [57–60]. However, other work has demonstrated that this may not be the case. In an extensive re-analysis of Kraiger and Ford's [60] military data, Black raters gave higher ratings to White ratees than to Black ratees [61]. Further analysis indicated that Black recruits consistently received lower ratings than White recruits from both Black and White raters [61]. Similarly, peer ratings in a sophomore-level engineering class demonstrated that minority students received lower ratings than non-minority students [49]. Finally, in some cases, no significant evidence of racial bias has been detected [62, 63]. Recently, a study of peer assessment in large general education classes taught using TBL found that while students of color contributed the same number of answers and suggestions as their peers, they received significantly lower peer evaluation scores than White students in three out of four assessment areas [21].

3. Study 1: Instructor Survey

3.1 Methods

Sixty-one instructors participated in the study, where 54 (38 women, 11 men, 5 no answer) completed the demographics section. Forty-six participants identified as White, three participants identified as Asian, and five participants identified as "Other." Forty-eight participants were native English speakers, while six were not. Participants spanned multiple colleges: Liberal Arts and Sciences (18), Agriculture and Life Sciences (9), Engineering (7), Design (7), Human Sciences (7), Business (5), and Veterinary Medicine (1).

Participants responded to the survey electronically. Instructors were asked about the frequency, type, and grading of peer assessment. They were asked for their perception of bias in peer evaluation and any training or other mitigations they implemented.

3.2 Results

3.2.1 Use of Peer Assessment

Instructors conducted surveys once (17), twice (14), three times (12), or more (13) per semester. Peer assessment was incorporated into student grades for 41 instructors. Over 75% (43/55) required qualitative feedback, typically requiring one statement on strengths and one on areas of improvement.

3.2.2 Perceptions of Bias

Of instructors, 47.3% perceived bias in peer evaluations due to gender (10 mentions), race (6), interpersonal relationships (6), language (3), and "gaming the system" (5). Selected quotes in Table 1 illustrate the range of issues. Comments on gender bias focused on biases women face and self-deprecation. Instructors noted that many students are uncomfortable giving constructive feedback based on cultural norms, interpersonal relationships, or a lack of compliance.

3.2.3 Mitigation of Bias

Of instructors, 51.9% said they had taken steps to mitigate bias. Common strategies include class discussion of appropriate evaluation criteria (10 mentions), the instructor "checking" the evaluations (7), and class discussion of sources of bias (5). Selected quotes from instructors are presented in Table 2. Instructors mitigated bias through team selection to balance and maximize diversity. Instructors also framed the evaluations as informative while emphasizing the purpose and content of good feedback.

4. Study 2: Student Survey

4.1 Participants

Data analysis was conducted on the 342 participants who completed more than half of the survey (203 women, 133 men, 5 no answer, 1 other). "Gender" was analyzed for two categories (Men and Women) due to the lack of participants in the "other" and "prefer not to say" categories. Participants were 18–55, with a median age of 20. Two-hundred ninety-five participants were white, 20

Table 1. Range of issues instructors mentioned in their replies

Type	Quote
Gender	"Females are more likely to deprecate their work in teams." "I have perceived sexist biases against women." "I've actually been very impressed by the maturity of the student responses but there are issues of women being evaluated differently than men that I've noticed." "I've noticed that the ways in which feedback is given varies based on gender – women tend to be less direct unless they are giving feedback anonymously"
Race, Ethnicity, Country of Origin	"Students of color have raised concerns that they are not being treated fairly" "I have had students from China tell me they are very uncomfortable giving a negative evaluation." "I have seen white students think that Chinese students were not taking the class seriously by not participating when in actuality the Chinese students were struggling with the language."
Interpersonal Relationships	". . . students not feeling comfortable saying negative things about their peers, especially bullies- so they receive good feedback." "Some members seem to give higher ratings to peers automatically so as not to hurt feelings or to 'be nice.' Conversely, some students who have personality conflicts with a peer seem to give ratings that are much lower than what other peers give the same student." "I've had issues of students grading assessments of their peers whom they dislike."
Non-Compliance/ "Gaming"	"Students don't report actual peer performance. They're biased against the feedback form." "I have found in both graduate and undergraduate courses that students tend to be very uncritical of their peers. They argue that they know how hard it is to do the work, so they should be praised for doing it rather than have a list of comments and a rating that shows where improvement is needed." "Students tend not to want to honestly peer assess their team members even if they complain that the student is not participating effectively in the team. I do not carry out peer assessment for that reason now."
Other	"Lower SES or non-traditional students receive lower grades."

Table 2. Selected quotes from instructor responses on methods of mitigating bias

Methods of Mitigating Bias – Selected Quotes
"For the qualitative responses, we do discuss that assessment should be balanced and informative. Specifically I discuss assessment as a source of information for the student being assessed rather than a positive or negative evaluation."
"Last spring, at the outset of class, I asked each team to anticipate the types of challenges they might encounter in the context of teams where I have aimed to maximize for diversity (e.g., gender, farm background or not, ethnicity, race, technology, sustainability, etc.) and asked each team to compose a team compact by consensus and then we share those statements and discuss as a whole class; before and after rounds of Peer Evals, I talk about what constitutes constructive comments"
"I've stated that peer assessments shouldn't be based on liking, but rather on actual work contributed."
"I discuss sources of potential bias and ask students to focus on the evaluation without bias."
"I try to set up my teams in order to minimize bias (make sure teams are gender balanced and that there is either zero or more than one person of color on teams to avoid someone being a 'token'."
"I use the output from CATME to identify outlier ratings, like those that would occur from personal conflicts or attempts to manipulate the ratings. When these kinds of behaviors are flagged in CATME, I follow-up by holding individual conferences with each student involved to better understand the ratings that were assigned."

Table 3. Student survey questions. All responses are Likert scale (1-Strongly Disagree; 7-Strongly Agree) unless otherwise indicated.

Question
I like working in teams on class projects.
I avoid classes that involve teamwork.
When working on a team in a class or for a class project, have you felt any biases from your teammates? (Yes/No/Maybe)
Follow-up: Please describe the bias(es) you have felt from your teammates (Free Response)
Follow-up: How frequently have you felt bias from your teammates? (Likert scale: 1-Never; 5-Always)
When working on a team in a class or on a class project, I feel respected by my teammates.
How many classes have you taken that use peer assessment? (number)
I feel that the peer assessments I have received were fair.
I feel that the peer assessments I have given were fair.
Have you felt any biases in peer assessments you have received? (Yes/No/Maybe)
Follow-up: Please describe the bias(es) you have felt in the peer evaluations you have received (Likert scale: 1-Never; 5-Always)
In what ways, if any, do you think bias could affect the peer assessments you receive from classmates? (Free Response)
In what ways, if any, do you think bias could affect the peer assessments you give classmates? (Free Response)

were Asian, 16 were Hispanic or Latinx, six were Black or African American, three were another race or ethnicity, and two were Indigenous American or Pacific Islander. The variable "race" was simplified into two categories: Students of Color (SOC) and White.

For 321 participants, English was their first language. For the 19 participants who did not initially speak English, the mean speaking time was 18.1 years (n = 18, SD = 4.1 years). Participant college affiliations were: Engineering (110), Liberal Arts and Sciences (75), Agriculture and Life Sciences (58), Human Sciences (30), Business (30), Design (27), Graduate (6), and Veterinary Medicine (6). All class levels were represented in the study: 66 freshmen, 63 sophomores, 67 juniors, 89 seniors, and 54 graduate students. Seventeen participants reported being international students, while 325 were not.

Table 4. Overall means and standard deviations of survey results, by question (N = 342).

Question	Mean (SD)
Like working in teams	4.1 (1.7)
Avoid classes that involve teamwork	3.2 (1.5)
Feel respected by teammates	5.1 (1.2)
Classmates want to be on a team with me	5.0 (1.3)
Peer assessments received were fair	5.5 (1.2)
Peer assessments given were fair	5.9 (1.1)

4.2 Procedure

Participants took the survey electronically. Participants were allowed to skip questions (summarized in Table 3).

4.3 Results

4.3.1 Overall Results

Means and standard deviations for each Likert survey question are reported in Table 4.

4.3.2 Differences by Gender, Race, and English Speaker Status

Women ($M = 3.9$, $SD = 1.7$, $N = 203$) liked working in teams significantly less than Men ($M = 4.5$, $SD = 1.6$, $N = 133$), $t(328) = -3.02$, $p < 0.001$, $d = 0.36$. Women ($M = 5.1$, $SD = 1.2$, $N = 203$) felt significantly less respected than Men ($M = 5.5$, $SD = 1.1$, $N = 133$), $t(328) = -3.19$, $p < 0.001$, $d = 0.35$. Women ($M = 5.4$, $SD = 1.2$, $N = 181$) reported significantly lower perceived fairness in assessments received than Men ($M = 5.7$, $SD = 1.1$, $N = 120$), $t(293) = -2.58$, $p = 0.005$, $d = 0.28$. There were no significant differences for the remaining items.

4.3.3 Difference by Race

White students (M = 4.0, SD = 1.7, N = 286) liked working on a team significantly less than POC students (M = 4.7, SD = 1.5, N = 42), $t(328)$ = –2.24, p = 0.013, d = 1.2. There were no significant differences by race for the remaining items.

4.3.4 Differences by English Speaker Status, International Student Status, and Class Level

Students whose first language was not English (M = 4.8, SD = 1.9, N = 19) perceived the assessments received less fair than native English speakers (M = 5.6, SD = 1.1, N = 282) $t(301)$ = 1.8, p = 0.04, d = 0.52. There were no significant differences by international student status for the remaining items.

There were no significant differences for any of the items by international student status.

There were no significant differences for any of the items by class level.

4.3.5 Occurrence of Bias

Ninety-three participants (27.3%) reported they had felt bias from their teammates, while 63 (18.5%) might have felt bias, and 185 (54.2%) had not felt bias. Commonly felt biases were due to gender (31 mentions), race (14), age (10), major (11), and interpersonal relationships (20). Other sources reported with less frequency were due to cultural differences, academic standing, and identification as LGBTQIA+. Of those students who had or might have experienced bias, the mean frequency of experiencing bias was 2.5 (SD = 0.8, N = 156).

4.3.6 Bias in Peer Assessments

Thirty-one participants (9.2%) reported they had felt bias in the peer assessments they had received, while 31 (9.2%) might have felt bias, and 275 (81.6%) had not felt bias. Commonly felt biases were due to gender (13 mentions), personality (8), interpersonal relationships (10), age (5), ''gaming the system''/obligation (11), and potential retaliation (5). Of those students who had or might have experienced bias in their peer assessments, the mean frequency of experiencing bias was 2.8 (SD = 0.9, N = 62).

4.3.7 How Bias Could Affect Peer Assessments Received

Seven participants felt unlikely to be the recipient of a biased peer assessment. For the rest of the participants, friendship status (42 mentions), ''gaming the system''/obligation (34), and personality (28) were the most commonly mentioned ways in which they could be the target of bias. Other items mentioned were gender, race, age, ability, and perceived fluency in the language of class instruction.

4.3.8 Student Perceptions of Bias When Rating Peers

Many participants (14) responded that they tried to be impartial or that bias would not affect the peer assessments they gave. For the rest of the participants, friendship status (41 mentions), ''gaming the system''/obligation (33), and personality (29) were the most commonly mentioned ways in which they could give biased assessments. Other items mentioned were gender, age, perceived ability, and language.

5. Study 3: Peer Assessment Data Analysis

5.1 Participants

The analysis included 3,885 students in 115 classrooms within the Thinkspace (Thinkspace, 2021) learning tool. This data was linked to demographic data provided by the Registrar using a key to protect individual confidentiality. The dataset assumed a gender binary and was almost equally divided between females (1,972) and males (1,913). Table 5 gives the breakdown of participants by race or ethnicity. It should be noted that the Office of the Registrar codes international students as ''International'' as opposed to racial/ethnic categories such as ''White'' or ''Asian.'' Students coded as ''International'' could be of any race or ethnicity.

Five-hundred ten students were first-generation college students, while 3,374 were not. English was the first language of 3,576 students, while 309 students initially spoke another language. Students averaged 25 years of age (range: 20–62). Two-hundred forty-six students were international students, while 3,639 were United States natives. The academic colleges represented were human sciences (899), engineering (889), liberal arts and sciences (851), business (300), design (140), interdisciplinary (18), and veterinary medicine (6).

Table 5. Race or ethnicity of students in the dataset

Race or Ethnicity	Count
White (not Hispanic)	2,964
International	246
Prefer not to indicate	185
Hispanic (Spanish American)	181
Asian	120
Black (not Hispanic)	98
Multiracial	77
American Indian or Alaskan Native	10
Native Hawaiian or Pacific Islander	3

Table 6. Means and standard deviations by gender of reviewer and reviewee. Levels not connected by the same letter are significantly different

Reviewer (given)	Reviewee (received)	Letter Report	Mean	SD
F	M	A	10.2	1.6
F	F	A	10.2	1.7
M	M	B	9.9	1.8
M	F	C	9.6	1.7

5.2 Peer Assessment Procedure

The peer assessments employed the Balance Method [64], where each student is given a set number of points to distribute among team members. Typically, students are given 10 points per student to distribute among their team members (not counting themselves). For example, if a team had five members, each student would have 40 points to split among their four team members. The number of points to distribute may vary with team size and instructor preference. Sometimes instructors add other limitations, such as requiring every student to receive a different score.

5.3 Data Analysis Procedure

The data contained 24,180 peer assessments completed between the 2013 Spring and 2018 Fall semesters. Blank or testing data was removed (1,123 rows). The peer assessment data was linked to demographic data using the reviewer and reviewee ID codes. The demographic data included gender, race, first language, major, college, semester GPA, and cumulative GPA. Rows with no demographic data were removed (547 rows). For instructors who used a point value per teammate other than 10, the peer assessment data were standardized to be on the same scale as the rest. The final dataset included 22,510 ratings given by 3,885 students in 115 classes.

For significant results, peer assessment scores were analyzed using a multiway ANOVA t-with Tukey post-hoc analysis. Specific analyses included the effect on peer assessment scores (given and received) based on gender, race, international status, and English as a first language. The same factors were analyzed for effect on GPA. A significance level of $\alpha = 0.05$ was used throughout. Effect size was determined using Cohen's *d* [65].

5.4 Results

5.4.1 Effect of Gender on Peer Assessment Score

The mean score *received* by males ($M = 10.1$, $SD = 0.8$) was significantly higher than the mean score received by females ($M = 9.8$, $SD = 0.8$), $F(1, 22{,}509) = 23.7$, $p < 0.001$, $d = 0.38$). The mean score *given* by females ($M = 10.2$, $SD = 0.9$) was significantly higher than the mean score given by males ($M = 9.7$, $SD = 0.9$), $F(1, 22{,}509) = 228.3$, $p < 0.001$, $d = 0.47$). There was also a significant interaction between the gender of the reviewer and the gender of the reviewee. The mean score given by females rating males was significantly higher than the mean score given by males rating males ($p < 0.001$, $d = 0.18$) and males rating females ($p < 0.001$, $d = 0.36$), $F(1, 22{,}509) = 16.3$, $p < 0.001$). There was no significant difference between the mean score given by females rating males ($p = 0.20$) and females to females ($p = 0.23$). The mean score given by males rating males was significantly lower than the mean score given by females rating males ($p < 0.001$, $d = 0.18$), and females rating females ($p < 0.001$, $d = 0.24$). The mean score given by males rating males was significantly higher than males rating females ($p < 0.001$, $d = 0.18$). These results for the interaction of reviewer and reviewer gender are summarized in Table 6.

5.4.2 Effect of Ethnicity on Peer Assessment Score

The mean score *received* by white students ($M = 10.1$, $SD = 1.9$) was significantly higher than the mean score received by students of color ($M = 9.7$, $SD = 2.0$, $p < 0.001$, $d = 0.22$) and international students ($M = 9.5$, $SD = 2.6$, $p < 0.001$, $d = 0.24$), $F(2, 22{,}507) = 36.2$, $p = < 0.001$. The mean score given by white students ($M = 9.6$, $SD = 2.0$) was significantly lower than the mean score given by international students ($M = 9.9$, $SD = 2.0$, $p < 0.001$, $d = 0.19$). There were no significant differences in the scores given by students of color and international or white students. Additionally, there were no significant differences in the scores received by students of color and international students. There was also a significant interaction between the ethnicity of the reviewer and the ethnicity of the reviewee. These results for the interaction of reviewer (given) and reviewee (received) ethnicity are summarized in Table 7 and Table 8.

5.4.3 Effect of International Student Status on Peer Assessment Score

The mean score *received* by domestic students ($M = 10.1$, $SD = 1.9$) was significantly higher than the mean score received by international students ($M = 9.5$, $SD = 2.6$), $p < 0.001$, $d = 0.31$). The mean score given by domestic students ($M = 9.7$, $SD = 2.0$) was

Table 7. Means and standard deviations by ethnicity of reviewer and reviewee. Levels not connected by the same letter are significantly different

Reviewer (given)	Reviewee (received)	Letter Report	Mean	SD
International	White	A	10.2	2.0
Students of color	White	A	10.1	1.8
White	White	A	10.1	1.9
International	Students of color	A B	9.9	2.1
International	International	A B	9.7	2.2
White	Students of color	B	9.6	2.0
Students of color	International	B C	9.6	1.5
Students of color	Students of color	B C	9.6	2.1
White	International	C	9.2	1.7

Table 8. Statistics for the interaction of reviewer and reviewee ethnicity. Means for the levels in the first column are significantly higher than means for the second column

Level (reviewer, reviewee)	Level (reviewer, reviewee)	*p*	*d*
International, White	White, International	< 0.001	0.53
Students of color, White	White, International	< 0.001	0.52
White, White	White, International	< 0.001	0.49
International, Students of color	White, International	0.002	0.36
International, White	Students of color, Students of color	< 0.001	0.34
International, White	Students of color, International	0.022	0.35
Students of color, White	Students of color, Students of color	< 0.001	0.32
International, White	White, Students of color	< 0.001	0.35
White, White	Students of color, Students of color	< 0.001	0.26
International, International	White, International	0.016	0.29
Students of color, White	Students of color, International	0.038	0.28
Students of color, White	White, Students of color	< 0.001	0.34
White, White	Students of color, International	0.039	0.27
White, White	White, Students of color	< 0.001	0.31
White, Students of color	White, International	< 0.001	0.35

significantly lower than the mean score given by international students (M = 9.9, SD = 2.1, p < 0.001, d = 0.18). There was also a significant interaction between the country of origin of the reviewer and the country of origin of the reviewee. The mean score given by domestic students rating international students was significantly lower than that of international students rating domestic students ($p < 0.001$, $d = 0.47$). The mean score given by domestic students rating domestic students (p < 0.001, d = 0.40) was significantly higher than international students rating

International students (p = 0.007, d = 0.24). There were no significant differences between mean score given by international students rating domestic students and domestic students rating domestic students (p = 0.33). There were no significant differences between mean score given by international students rating domestic students and international students rating international students (p = 0.20). These results for the interaction of reviewer and reviewee international student status are summarized in Table 9.

5.4.4 Effect of English Speaker Status on Peer Assessment Score

The mean score received by native English speakers (M = 10.1, SD = 1.9) was significantly higher than the mean score received by non-native English speakers (M = 9.5, SD = 2.2, p < 0.001, d =

Table 9. Means and standard deviations by international student status of reviewer and reviewee. Levels not connected by the same letter are significantly different

Reviewer (given)	Reviewee (received)	Letter Report	Mean	SD
International	Domestic	A	10.1	1.9
Domestic	Domestic	A	10.0	2.0
International	International	A	9.7	2.0
Domestic	International	B	9.3	2.2

Table 10. Means and standard deviations by English speaker status of reviewer and reviewee. Levels not connected by the same letter are significantly different

Reviewer (given)	Reviewee (received)	Letter Report	Mean	SD
Non-native	Native	A	10.1	2.0
Native	Native	A	10.0	1.9
Non-native	Non-native	A	9.7	2.2
Native	Non-native	B	9.3	2.1

0.28). The mean score given by native English speakers (M = 9.7, SD = 2.1) was significantly lower than the mean score given by non-native English speakers (M = 9.9, SD = 1.8, $p < 0.001$, d = 0.13). There was also a significant interaction between the country of origin of the reviewer and the country of origin of the reviewee. The mean score given by native English speakers to non-native English-speaking students (M = 9.3, SD = 2.1) was significantly lower than the mean score given by non-native English speakers rating English speakers (M = 10.1, SD = 2.0, $p < 0.001$, d = 0.41), native English speakers rating native English speakers (M = 10.0, SD = 1.9, $p < 0.001$, d = 0.38), and non-native English speakers rating non-native English speakers (M = 9.7, SD = 2.0, $p < 0.032$, d = 0.21). There were no significant differences between mean score given and received in any other pairs. These results for the interaction of reviewer and reviewee language are summarized in Table 10.

5.4.5 Effect of Demographics on GPA

To determine if differences in GPA could explain the differences in peer assessment score by gender, international student status, and native language, we analyzed the effect of these demographic variables on GPA. The mean GPA for female students (M = 3.14, SD = 0.59) was significantly higher than the mean GPA for male students (M = 2.93, SD = 0.67). The mean GPA for international students (M = 3.28, SD = 0.68) was significantly higher than the mean GPA for domestic students (M = 3.03, SD = 0.63), $F(1, 22{,}509) = 227.5$, $p < 0.001$, d = 0.38. The mean GPA for students whose first language was not English (M = 3.12, SD = 0.69) was significantly higher than the mean GPA for students whose first language was English (M = 3.03, SD = 0.63), $F(1, 22{,}509) = 21.6$, $p < 0.001$, d = 0.14.

There was a significant main effect of race on GPA, $F(2, 22{,}508) = 258.3$, $p < 0.001$). The mean GPA for international students (M =3.28, SD = 0.67) was significantly higher than the mean GPA for white students (M =3.05, SD = 0.62, $p < 0.001$, d = 0.36) and students of color (M = 2.84, SD = 0.64, $p < 0.001$, d = 0.67). The mean GPA for white students (M = 3.05, SD = 0.62) was significantly higher than the mean GPA for students of color (M = 2.84, SD = 0.64, $p < 0.001$, d = 0.33). These results are summarized in Table 11.

Table 11. Means and standard deviations of GPA by race. Levels not connected by the same letter are significantly different

Race	Letter Report	Mean	SD
International	A	3.28	0.67
White	B	3.05	0.62
Students of color	C	2.84	0.64

6. Discussion

The three studies reported above established evidence of bias in peer assessment. In the student survey, participants noted bias in the assessments they had received and reflected on how the peer assessments they gave could be biased. Similarly, instructors again noted bias in their classrooms and assessments. Finally, the evidence of bias is shown in peer assessment scores themselves, which cannot be fully explained by student achievement (e.g., GPA).

6.1 Perception of Bias in the Classroom

Study 1 indicated that many instructors (47%) have perceived bias in peer evaluations. These results are even higher than those reported in 2009 [44], where 27% of professors and 25% of graduate instructors noticed bias in their classrooms. The types of biases observed in both studies are similar, with race, gender, and country of origin being represented. These perceptions are given further credibility by the results of study 3, which show that gender, race, and country of origin can have a negative effect on peer evaluation scores. While many of the biases described by the instructors in this study were what often comes to mind when reading the word ''bias'' (e.g., racism, sexism), others were specific to collaborative learning environments and peer assessment, such as a general unwillingness to give negative ratings or feedback. This unwillingness to be critical of peers has been observed in other studies of peer assessment [66, 67].

6.2 Student Perceptions of Peer Assessment

Female students reported significantly lower enjoyment of working in teams, respect from teammates,

and perceptions of fairness in their peer assessments than men. Similarly, 66% of students who had experienced bias in a classroom team and 60% of students who had experienced bias in peer assessments were women. These findings are consistent with work showing that male students report more positive attitudes about peer assessment than female students [33, 53]. The results of the third study demonstrate that these lower perceptions of fairness reflect the reality of the peer assessment scores women earn.

Student participants made many of the same observations as instructors in terms of bias in peer assessment due to personality and a reluctance to give negative feedback. Descriptions of "personality" bias again follow the general descriptions of social style, with those who exhibit the expressive style often being portrayed as receiving higher ratings. However, many students mentioned bias due to friendship, a commonly noted potential source of biased peer assessments [68, 69].

In addition to the biases they observed, students were asked how bias *could* affect assessments given and received. Respondents often focused on the influence of personality and friendship. However, students noted categorical biases (e.g., gender, language) that could influence how they rate and are rated.

This indicates that these students are aware of the potential for bias, which is an important first step in mitigation [70]. When reporting on biases they had perceived against themselves and how peer assessments *could* be biased, students mentioned the interconnected issues of retaliation and a general unwillingness to give negative feedback. The attention paid to these issues by participants shows that some students are coming into the peer assessment process already affected by prior experiences. These prior experiences may unwittingly lead students to rate their peers less honestly than they otherwise would.

6.3 Fairness of Rating by Gender

In Study 3, female raters received lower scores overall yet gave higher scores, which is similar to some previous findings [50, 51]. This contrasts with May and colleagues [71], who found that men received lower ratings overall. The finding for men receiving lower overall ratings is often explained by women earning higher GPAs (e.g., [24]). In the current study, females received lower peer assessment scores, and GPA does not fully explain this discrepancy. Unlike the studies referenced here [24, 25, 71], however, the current work analyzed a large body of data outside the realm of a single classroom or department. Further, female raters showed no significant differences in their ratings of males or females. This finding could suggest that the female raters were fairer in terms of gender.

6.4 Fairness of Rating by International Status and Race

While the effect of gender on peer assessment scores has received a considerable amount of study, international student status has received much less. International student enrollment in the United States has greatly increased in the past twenty years; however, these students are often the targets of nativism, racism, and other forms of discrimination [72, 73]. In the analysis of peer assessment scores, domestic students gave international students lower peer assessment scores than they gave their domestic-based peers. Previous work has demonstrated that international students are more generous in their peer assessment ratings than domestic students [74]. When race was added to the analysis, the outcomes remained the same. White domestic students received higher scores than international students of any race. Conversely, white domestic students rate international students lower than domestic white students. In both instances, international students had higher GPAs than white and domestic-based students, so their lower peer assessment scores are less likely to explain differences in achievement.

White students received higher scores than students of color. Conversely, white students rated students of color lower than other white students. These results show that the contributions of students of color are less valued and are similar to previous findings [21]. Unlike the comparisons of peer assessment scores with GPA based on gender, the students of color in this study had lower peer assessment scores and GPAs. However, numerous interacting factors specific to the experiences of students of color may affect GPA. Students of color, specifically Black, Latinx, and Native American students, historically earn lower GPAs than their white counterparts [75, 76]. Students from these groups are disproportionally working learners from low-income backgrounds [77]. This suggests that since students of color are more likely to come from lower-income backgrounds and work longer hours while in college, their GPAs suffer. Minority students often attend high schools with lower instructional quality than white students, which is a factor in differences in college achievement [75]. As a result, the lower peer assessment scores received by students of color cannot be entirely separated from the interconnected factors of bias, student team performance, and GPA.

For women and international students, peer assessment scores and GPA moved in opposite

directions. Conversely, for students of color, peer assessment scores and GPA moved in similar directions. When interpreting the results of these studies, it should be noted that GPA is not suggested to be a direct predictor of peer assessment score. Some work has shown that GPA and peer assessment scores correlate with students earning higher GPAs and higher peer assessment scores; however, the scope of this work is limited to small sample sizes [78]. The present analysis of GPA and peer assessment scores was conducted because GPA has been suggested to be a potential factor in differing peer assessment scores [24]. Nevertheless, it is possible that other factors (e.g., student team performance) affect peer assessment scores more than GPA or bias.

6.5 Implications for Educators

The three studies have established the need for further work to address potential biases in peer assessment. Promoting awareness among students of the potential for bias in peer assessment would be a first step to creating fairer assessments. However, awareness alone may not be sufficient. Bias reduction training should focus both on awareness of bias as well as corrective actions [79]. This could be in the form of training or active educational experiences. Such training would need to be grounded in understanding the underlying behaviors that reinforce biased stereotypes (whether recognized or not). More work is necessary to establish the requirements of peer assessment fairness training interventions.

Another implication is the need for some interventions to address bias in peer assessment. Previous work has found that faculty discomfort in addressing non-disciplinary topics is a significant challenge in addressing equity [80]. Faculty need tools and strategies to address these issues. This may include how teams are formed as well as how to train students to provide fairer assessments.

6.6 Limitations

The studies were limited by the demographics of the institution in which they were deployed. As most participants were White and from the United States, the results of these studies are not generalizable to all students or instructors, especially students of color, international students, and their instructors. The student data coding of ''International'' is not a race or ethnicity. Using this code instead of the actual race or ethnicity of the international student makes it difficult to fully understand the issue of bias for students of color. Finally, any assessment method should be evaluated for fairness and its capability to assess diverse groups fairly [81].

7. Conclusion

The use of peer assessment in the classroom is associated with better learning outcomes for students. However, the potential for bias in peer assessment threatens these outcomes. These studies examined the issue of biased peer assessments from multiple perspectives. The results of the survey studies revealed that both students and faculty perceive similar biases in peer assessments. Additionally, the analysis of over 20,000 peer assessment ratings revealed significant differences in peer assessment scores received based on gender, race or ethnicity, international student status, and native language. These differences also extended to the scores raters gave to various groups (e.g., female students rating male students or female students rating female students) and are not fully explained by differences in achievement (e.g., GPA).

Future work in this area could conduct these types of analyses across a more comprehensive, more diverse set of educational institutions. While more work needs to be done to explore the issue of bias, it is also essential to work toward a solution. Therefore, an important future direction for scholarship within engineering education is using this work as a basis for creating peer assessment rater training focused on fairness [83].

8. Authors Positionality Statement

As Holmes [82] notes, researchers should assess how their positions and experiences might contribute to their interpretations of the experiences of others. Because this project involves interpreting data relating to biases the authors may not have experienced, it is important to state their positionality. The primary author is a middle-class US-born White woman with graduate research training in industrial engineering.

The secondary author is a US-born White man and a first-generation US Citizen. The third member of the team is also a US-born White woman. All three members of the team identify as cis-gendered. All three authors have experience teaching undergraduate students, and two have taught graduate students. All have employed peer assessment in their classrooms. The research team used reflexive practices during the research and this positionality statement to make transparent our relationship to the issues and data presented here and to mitigate any potential limitations created by that relationship. We do not seek to remove ourselves from our positionality but to acknowledge its presence within the work presented here.

Acknowledgements – The authors would like to thank Dr. Cassandra Dorius for early advice on data analysis.

References

1. L. K. Michaelsen, A. B. Knight and L. D. Fink, *Team-Based Learning: A Transformative Use of Small Groups*. Sterling, VA: Stylus Publishing, 2004.
2. L. K. Michaelsen, M. Sweet and D. X. Parmelee, Team-Based Learning: Small Group Learning's Next Big Step, *New Dir. Teach. Learn.*, **103**(116), 2011.
3. R. E. Allen, J. Copeland, A. S. Franks, R. Karimi, M. McCollum, D. J. Riese II and A. Y. F. Lin, Team-Based Learning in US Colleges and Schools of Pharmacy, *Am. J. Pharm. Educ.*, **77**(6), pp. 1–9, Aug. 2013.
4. S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, in *Proceedings of the National Academy of Sciences of the United States of America*, **111**(23), pp. 8410–8415, 2014.
5. R. M. Lima, P. H. Andersson and E. Saalman, Active Learning in Engineering Education:(re)introduction, *Assess. Eval. High. Educ.*, **42**(1), pp. 1–4, Jan. 2016.
6. D. Beneroso and M. Erans, Team-based learning: an ethnicity-focused study on the perceptions of teamwork abilities of engineering students, *Eur. J. Eng. Educ.*, **46**(5), pp. 678–689, 2020.
7. R. E. Slavin and E. Oickle, Effects of Cooperative Learning Teams on Student Achievement and Race Relations: Treatment by Race Interactions, *Sociol. Educ.*, **54**(3), p. 174, Jul. 1981.
8. L. Springer, M. E. Stanne and S. S. Donovan, Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis, *Rev. Educ. Res.*, **69**(1), pp. 21–51, Mar. 1999.
9. H. S. Kim, Uncertainty analysis for peer assessment: oral presentation skills for final year project, *J. Eng. Educ.*, **39**(1), pp. 68–83, 2014.
10. P. M. Sadler and E. Good, The impact of self- And peer-grading on student learning, *Educ. Assess.*, **11**(1), pp. 1–31, 2006.
11. K. J. Topping, Peers as a source of formative and summative assessment, in *SAGE handbook of research on Classroom Assessment*, SAGE Publications, pp. 395–412, 2013.
12. M. Barak and S. Rafaeli, On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning, *Int. J. Hum. Comput. Stud.*, **61**(1), pp. 84–103, 2004.
13. P. Dawson, Assessment rubrics: towards clearer and more replicable design, research and practice, *Assess. Eval. High. Educ.*, **42**(3), pp. 347–360, 2015.
14. C. M. Cestone, R. E. Levine and D. R. Lane, Peer assessment and evaluation in team-based learning, *New Dir. Teach. Learn.*, **2008**(116), pp. 69–78, 2008.
15. T. Brown, J. Rongerude, B. Leonard and L. C. Merrick, Best practices for online team-based learning: Strengthening teams through formative peer evaluation, *New Dir. Teach. Learn.*, **2021**(165), pp. 53–64, 2021.
16. G. Thondhlana and D. Z. Belluigi, Students' reception of peer assessment of group-work contributions: problematics in terms of race and gender emerging from a South African case study, *Assess. Eval. High. Educ.*, **42**(7), pp. 1118–1131, 2016.
17. A. Planas Lladó, L. F. Soley, R. M. Fraguell Sansbelló, G. A. Pujolras, J. P. Planella, N. Roura-Pascual, J. J. Suñol Martínez and L. M. Moreno, Student perceptions of peer assessment: an interdisciplinary study, *Assess. Eval. High. Educ.*, **39**(5), pp. 592–610, 2014.
18. D. Moorman and D. Wicks-Smith, Poverty discrimination revealed through student peer evaluations, *Coll. Stud. J.*, **46**(1), pp. 141–149, 2012.
19. G. A. Boysen and D. L. Vogel, Bias in the Classroom: Types, Frequencies, and Responses, *Teach. Psychol.*, **36**(1), pp. 12–17, 2009.
20. R. Tucker, Sex does not matter: gender bias and gender differences in peer assessments of contributions to group work, *Assess. Eval. High. Educ.*, **39**(3), pp. 293–309, 2014.
21. C. Wayland, L. Walker and A. Ferrara, Opening the Black Box: How Gender and Race Mediate Collaborative Learning in the Classroom, in *Southeastern Women's Studies Association Annual Meeting*, 2014.
22. R. Gunter and A. Stambach, Differences in Men and Women Scientists' Perceptions of Workplace Climate, *J. Women Minor. Sci. Eng.*, **11**(1), pp. 97–116, 2005.
23. R. Baltà Salvador, M. Peña Carrera, N. Olmedo Torre and R. Wang, Factors influencing career choice, perceived discrimination, and segregation of foreign-born engineering undergraduates, *Int. J. Eng. Educ.*, **39**(3), pp. 570–583, 2023.
24. G. L. May and L. E. Gueldenzoph, The Effect of Social Style on Peer Evaluation Ratings in Project Teams, *J. Bus. Commun.*, **43**(1), pp. 4–20, 2006.
25. M. Espey, Gender and peer evaluations, *J. Econ. Educ.*, **53**(1), pp. 1–10, 2021.
26. N. Falchikov and J. Goldfinch, Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks, *Rev. Educ. Res.*, **70**(3), pp. 287–322, 2000.
27. S. J. Hanrahan and G. Isaacs, Assessing Self- and Peer-assessment: The students' views, *High. Educ. Res. Dev.*, **21**(1), pp. 53–70, 2010.
28. S. Miedijensky and T. Tal, Embedded Assessment in Project-based Science Courses for the Gifted: Insights to inform teaching all students, *Int. J. Sci. Educ.*, **31**(18), pp. 2411–2435, 2008.
29. S. J. Deeley and C. Bovill, Staff student partnership in assessment: enhancing assessment literacy through democratic practices, *Assess. Eval. High. Educ.*, **42**(3), pp. 463–477, 2015.
30. L. Li and F. Gao, The effect of peer assessment on project performance of students at different learning levels, *Assess. Eval. High. Educ.*, **41**(6), pp. 885–900, 2015.
31. S. Bloxham and A. West, Understanding the rules of the game: marking peer assessment as a medium for developing students' conceptions of assessment, *Assess. Eval. High. Educ.*, **29**(6), pp. 721–733, 2010.
32. L. A. J. Stefani, Peer, self and tutor assessment: Relative reliabilities, *Stud. High. Educ.*, **19**(1), pp. 69–75, 2006.
33. M. L. Wen and C. C. Tsai, University students' perceptions of and attitudes toward (online) peer assessment, *High. Educ.*, **51**(1), pp. 27–44, 2006.
34. K. Topping, Peer Assessment Between Students in Colleges and Universities, *Rev. Educ. Res.*, **68**(3), pp. 249–276, 1998.
35. E. Walsh, M. Neelen and E. Kapros, Competency Analytics in the Workplace through Continuous Peer Feedback, in *Proceedings – IEEE 17th International Conference on Advanced Learning Technologies, ICALT 2017*, pp. 224–226, 2017.

36. ABET, Criteria for Accrediting Engineering Technology Programs, 2020–2021. [Online]. Available: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-technology-programs-2020-2021/. [Accessed: 20-Jul-2023].
37. D. Magin and P. Helmore, Peer and Teacher Assessments of Oral Presentation Skills: How reliable are they?, *Stud. High. Educ.*, **26**(3), pp. 287–298, 2010.
38. E. Samuel, Racism in Peer-Group Interactions: South Asian Students' Experiences in Canadian Academe, *J. Coll. Stud. Dev.*, **45**(4), pp. 407–424, 2004.
39. D. Dancer and P. Kamvounias, Student involvement in assessment: a project designed to assess class participation fairly and reliably, *Assess. Eval. High. Educ.*, **30**(4), pp. 445–454, 2010.
40. V. Aryadoust, Gender and Academic Major Bias in Peer Assessment of Oral Presentations, *Lang. Assess. Q.*, **13**(1), pp. 1–24, 2016.
41. J. F. Dovidio, K. Kawakami and S. L. Gaertner, Implicit and explicit prejudice and interracial interaction, *J. Pers. Soc. Psychol.*, **82**(1), pp. 62–68, 2002.
42. A. M. La Greca, S. Lai, B., Chan and W. Herge, Peer Assessment Strategies, in *Diagnostic and Behavioral Assessment in Children and Adolescents: A Clinical Guide*, B. D. McLeod, A. J. Dpcc, and T. H. Ollendick, Eds. New York, NY, USA: The Guilford Press, pp. 417–488, 2013.
43. J. F. Dovidio and R. H. Fazio, New technologies for the direct and indirect assessment of attitudes, in *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, J. M. Tanur, Ed. Washington, DC: Russell Sage Foundation, pp. 204–237, 1992.
44. G. A. Boysen, D. L. Vogel, M. A. Cope and A. Hubbard, Incidents of Bias in College Classrooms: Instructor and Student Perceptions, *J. Divers. High. Educ.*, **2**(4), pp. 219–231, 2009.
45. P. Aeby, R. Fong, M. Vukmirovic, S. Isaac and R. Tormey, The Impact of Gender on Engineering Students' Group Work Experiences, *Int. J. Eng. Educ.*, **35**(3), pp. 756–765, 2019.
46. C. Brindley and S. Scoffield, Peer Assessment in Undergraduate Programmes, *Teach. High. Educ.*, **3**(1), pp. 79–90, 2006.
47. S. Thiam, T. Madruga, A. Vasquez, R. Kothari and G. G. Krauss, Is it Really a Choice Between Quantity and Quality for Peer Feedback?, *Int. J. Eng. Educ.*, **36**(2), pp. 687–701, 2020.
48. F. Pajares and M. J. Johnson, Self-efficacy beliefs and the writing performance of entering high school students – Pajares – 1996 – Psychology in the Schools – Wiley Online Library, *Psychol. Sch.*, **33**(2), pp. 163–175, 1996.
49. D. B. Kaufman, R. M. Felder and H. Fuller, Accounting for Individual Effort in Cooperative Learning Teams, *J. Eng. Educ.*, **89**(2), pp. 133–140, Apr. 2000.
50. R. E. Bryan, A. J. Krych, S. W. Carmichael, T. R. Viggiano and W. Pawlina, Assessing Professionalism in Early Medical Education: Experience with Peer Evaluation and Self-Evaluation in the Gross Anatomy Course, *Ann. Med. Singapore*, **34**(8), p. 486, 2005.
51. W. R. Sherrard, F. Raafat & Richard and R. R. Weaver, An Empirical Study of Peer Bias in Evaluations: Students Rating Students, *J. Educ. Bus.*, **70**(1), pp. 43–47, 1994.
52. T. Gatfield, Examining Student Satisfaction with Group Projects and Peer Assessment, *Assess. Eval. High. Educ.*, **24**(4), pp. 365–377, 2006.
53. K. J. Topping, Methodological quandaries in studying process and outcomes in peer assessment, *Learn. Instr.*, **20**(4), pp. 339–343, 2010.
54. C. Rees, Self-assessment scores and gender, *Med. Educ.*, **37**(6), pp. 572–573, 2003.
55. D. S. Lind, S. Rekkas, V. Bui, T. Lam, E. Beierle and E. M. Copeland, Competency-Based Student Self-Assessment on a Surgery Rotation, *J. Surg. Res.*, **105**(1), pp. 31–34, 2002.
56. M. Das, D. Mpofu, E. Dunn and J. H. Lanphear, Self and tutor evaluations in problem-based learning tutorials: is there a relationship?, *Med. Educ.*, **32**(4), pp. 411–418, 1998.
57. J. A. Cox and J. D. Krumboltz, Racial Bias in Peer Ratings of Basic Airmen, *Sociometry*, **21**(4), p. 292, 1958.
58. J. E. Dejung and H. Kaplan, Some differential effects of race of rater and ratee on early peer ratings of combat aptitude, *J. Appl. Psychol.*, **46**(5), pp. 370–374, 1962.
59. F. J. Landy and J. L. Farr, Performance rating, *Psychol. Bull.*, **87**(1), pp. 72–107, 1980.
60. K. Kraiger and J. K. Ford, A Meta-Analysis of Ratee Race Effects in Performance Ratings, *J. Appl. Psychol.*, **70**(1), pp. 56–65, 1985.
61. P. R. Sackett and C. L. Z. DuBois, Rater-Ratee Race Effects on Performance Evaluation: Challenging Meta-Analytic Conclusions, *J. Appl. Psychol.*, **76**(6), pp. 873–877, 1991.
62. F. L. Schmidt and R. H. Johnson, Effect of Race on Peer Ratings in an Industrial Situation, *J. Appl. Psychol.*, **57**(3), pp. 237–241, 1973.
63. E. D. Pulakos, L. A. White, S. H. Oppler and W. C. Borman, Examination of Race and Sex Effects On Performance Ratings, *J. Appl. Psychol.*, **74**(5), pp. 770–780, 1989.
64. L. K. Michaelsen, L. D. Fink and A. Knight, Designing Effective Group Activities: Lessons for Classroom Teaching and Faculty Development, *To Improv. Acad.*, **16**(1), pp. 373–397, 1997.
65. J. Cohen, *Statistical Power Analysis for the Behavioural Sciences*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc, 1988.
66. K. J. Topping, Peer Assessment, *Theory Pract.*, **48**(1), pp. 20–27, 2009.
67. L. Arnold, C. K. Shue, B. Kritt, S. Ginsburg and D. T. Stern, Medical students' views on peer assessment of professionalism, *J. Gen. Intern. Med.*, **20**(9), pp. 819–824, 2005.
68. S. Brown and P. Knight, *Assessing Learners in Higher Education*, London, UK: Routledge, 1994.
69. A. M. Langan, C. P. Wheater, E. M. Shaw, B. J. Haines, W. R. Cullen, J. C. Boyle, D. Penney, J. A. Oldekop, C. Ashcroft, L. Lockey and R. F. Preziosi, Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of assessment criteria, *Assess. Eval. High. Educ.*, **30**(1), pp. 21–34, 2010.
70. P. G. Devine, P. S. Forscher, A. J. Austin and W. T. L. Cox, Long-term reduction in implicit race bias: A prejudice habit-breaking intervention, *J. Exp. Soc. Psychol.*, **48**(6), pp. 1267–1278, 2012.
71. G. L. May, The Effect of Rater Training on Reducing Social Style Bias in Peer Evaluation, *Bus. Commun. Q.*, **71**(3), pp. 297–313, 2008.
72. C. W. Yao, C. A. George Mwangi and V. K. Malaney Brown, Exploring the intersection of transnationalism and critical race theory: a critical race analysis of international student experiences in the United States, *Race Ethn. Educ.*, **22**(1), pp. 38–58, 2018.
73. J. J. Lee and C. Rice, Welcome to America? International student perceptions of discrimination, *High. Educ.*, **53**(3), pp. 381–409, 2007.

74. Y. Cinar and A. Bilgin, Peer assessment for undergraduate teamwork projects in petroleum engineering, *Int. J. Eng. Educ.*, **27**(2), pp. 310–322, 2011.
75. J. Fletcher and M. Tienda, Race and Ethnic Differences in College Achievement: Does High School Attended Matter?, *Ann. Am. Acad. Pol. Soc. Sci.*, **627**(1), pp. 144–166, 2010.
76. J. Woo, C. Green and M. Matthews, *Profile of 2007–8 first-time bachelor's degree recipients in 2009*, Washington DC, 2013.
77. A. P. Carnevale and N. Smith, Balancing Work and Learning: Implications for Low-Income Students, *New Dir. Teach. Learn.*, **2018**(116), pp. 69–78, 2018.
78. N. Al Mortadi, S. S. Al-Houry, K. H. Alzoubi and O. F. Khabour, Effectiveness of Peer Evaluation in Learning Process: A Case from Dental Technology Students, *Open Dent. J.*, **14**(1), pp. 692–697, 2020.
79. E. R. Carter, I. N. Onyeador and J. Neil A. Lewis, Developing & delivering effective anti-bias training: Challenges & recommendations, *Soc. Work Soc. Policy*, **6**(1), 2020.
80. G. Pfeifer and E. A. Stoddard, Equitable and Effective Student Teams: Creating and Managing Team Dynamics for Equitable Learning Outcomes, in *Project-Based Learning in the First Year: Beyond All Expectations*, K. Wobbe and E. A. Stoddard, Eds. Routledge, pp. 220–263, 2023.
81. K. A. Douglas, A. Rynearson, E. Purzer and J. Strobel, Reliability, Validity, and Fairness: A Content Analysis of Assessment Development Publications in Major Engineering Education Journals, *Int. J. Eng. Educ.*, **32**(5(A)), pp. 1960–1971, 2016.
82. G. A. D. Holmes, Researcher Positionality – A Consideration of Its Influence and Place in Qualitative Research – A New Researcher Guide, *Shanlax Int. J. Educ.*, **8**(4), pp. 1–10, 2020.
83. J. H. Stonewall, M. C. Dorneich and J. Rongerude, Training to Increase Student Perceptions of Fairness in Peer Assessment, *Assess. Eval. High. Educ.*, pp. 1–17, 2024.

Jacklin H. Stonewall, PhD is a Senior UX Researcher at John Deere, Inc. She earned her PhD in the Departments of Industrial and Manufacturing Systems Engineering and Human Computer Interaction at Iowa State University. Jacklin obtained her BS and MS from Iowa State University in Aerospace Engineering and Industrial Engineering, respectively. Her research focuses on creating equitable systems that not only allow but encourage use by all individuals, regardless of ability. She is a UX expert who has lead the usability evaluation of electronic voting systems for disabled users, evaluation of the gender perceptions of website design, and usability evaluations of multimodal, virtual /extended reality traffic simulation environments.

Michael C. Dorneich, PhD is the Joseph Walkup Professor of Engineering at Iowa State University in the Industrial and Manufacturing Systems Engineering Department. He earned his PhD in Industrial Engineering in human factors at the University of Illinois at Urbana-Champaign. Dr. Dorneich's research interests focus on human factors, human-computer interaction, intelligent tutoring systems, human-autonomy teaming, and team-based learning. His research includes work in how to increase female participation in STEM, reduce bias in peer assessments, and apply team-based learning (TBL) techniques. His teaching methods include team projects and the application of team-based learning methods into the classroom. He was Chair of the Board of Education for the Twin Cities German Immersion School, a K-8 charter school in St. Paul, MN. He was the Faculty Coordinator of the Team-based Learning (TBL) Faculty Learning Community, with over 180 members. He has authored over 220 publications and holds 28 US and international patents.

Jane Rongerude, PhD is an associate professor in the Department of Community and Regional Planning at Iowa State University where she teaches classes in planning theory, housing, community planning, and planning and social justice. She earned her MCP and PhD in city and regional planning from the University of California at Berkeley. Her research investigates how housing policy and planning practices shape and maintain poverty places, and how they contribute to or inhibit opportunities for community transformation. Her research related to planning education focuses on pedagogical innovations related to Team Based Learning, and specifically the question of how to integrate diverse voices within the classroom. She currently serves on the executive committee for Inclusion, the LGBTQIA interest group for the Association of Collegiate Schools of planning and on the Diversity Committee for the Iowa Chapter of the American Planning Association.