A Review of Bias in Peer Assessment

Jacklin Hope Stonewall, Iowa State University

Jacklin Stonewall is a Ph.D. student in the Departments of Industrial and Manufacturing Systems Engineering and Human Computer Interaction at Iowa State University. Her research interests include: gender HCI, decision support systems, sustainability, and the creation of equitable cities and classrooms.

Prof. Michael Dorneich, Iowa State University

Dr. Michael C. Dorneich is an Associate Professor at Iowa State University in the Industrial and Manufacturing Systems Engineering Department. He graduated in 1999 from the University of Illinois at Urbana-Champaign with a PhD. in Industrial Engineering in the Human Factors Program. His research interests focus on human factors, human-computer interaction, and adaptive systems that enable people to be effective in the complex and often stressful environments found in aviation, military, robotic, and space applications. His teaching methods include team projects and the application of team-based learning methods into the classroom.

Cassandra Dorius

Jane Rongerude PhD, Department of Community and Regional Planning, Iowa State University

Jane Rongerude is an assistant professor in the Department of Community and Regional Planning at Iowa State University, where she teaches courses in planning theory, community and urban revitalization, housing, and neighborhood planning. Her research interests include the potential of planning as a tool to interrupt local system of poverty management and pedagogical innovations related to Team Based Learning, specifically how to integrate diverse voices within communities and classrooms. She earned her MCP and PhD in city and regional planning from the University of California at Berkeley.
A Review of Bias in Peer Assessment

Jacklin Stonewall¹, Michael Dorneich¹, PhD, Jane Rongerude², PhD, Cassandra Dorius³, PhD
¹Industrial and Manufacturing Systems Engineering, Iowa State University
²Community and Regional Planning, Iowa State University
³Human Development and Family Studies, Iowa State University

The use of teams and team-centric pedagogies such as Team Based Learning (TBL) in classrooms has been shown to increase engagement and lead to better overall learning outcomes. Active learning pedagogies such as TBL are also promoted as promising strategies for engaging underrepresented students. For many instructors, especially those using TBL, peer assessments are integral to the classroom environment as tools for both monitoring team performance and ensuring accountability. However, concerns have developed regarding the fairness of peer assessments due to student biases. Research on TBL classrooms finds that women and students of color do not have the same experiences as their white male counterparts. Additionally, bias has been observed in peer assessment scores with respect to race, gender, and socioeconomic status. As more instructors recognize the benefits of teams and integrate them into their classes, the use of peer assessments also increases, highlighting the need for a fair peer assessment process. Through literature review, this work will present an initial description of the issues involved and identify the extent to which bias has been observed to affect peer assessment scores. This is part of a larger project that aims to use this information to design and evaluate fairer peer assessment processes.

INTRODUCTION

Biases, both implicit and explicit, negatively impact the way people perceive members of disadvantaged groups. Unfortunately, biases often extend into the classroom environment [1], [2]. In one academic year, 38% of professors surveyed perceived an act of bias in their classes [2]. In small group learning classrooms, these biases have been shown to manifest in many ways, including in peer assessments [3].

The employment of small group learning strategies (such as cooperative learning or Team Based Learning) in classroom environments has been shown to increase student achievement, attendance, engagement, and lead to better overall learning outcomes [4], [5], [6]. Because of these outcomes, team-based pedagogies and cooperative learning practices have been incorporated on college campuses as a strategy to improve the classroom engagement of underrepresented students. Indeed, research shows that learning in teams positively affects objective outcomes (such as exam scores) for minority students [7], [8]. In many group learning classrooms, peer assessments are integral.

One of the most widely used team-centric pedagogies is team-based learning (TBL). The prevalence of TBL in the United States has been steadily rising, especially in the medical field [6]. In TBL classrooms, permanent teams are formed to maximize heterogeneity. TBL functions on four essential principles: 1. Properly formed teams that remain together for the duration of the term, 2. Readiness assurance assessments of individual’s and team’s pre-class preparation, 3. In-class, team application exercises that promote learning of material and team development, and 4. Peer assessments designed to monitor team performance, hold individuals accountable for their effort and contribution, and serve to improve overall team functioning [5]. Outside of specific pedagogies and classroom structures, peer assessments themselves have been shown to have far-reaching benefits. These assessments have been credited with empowering learners to engage more fully in the class [9] and increasing interactions among students and between students and instructors [10]. Assessments can foster the development of autonomy and maturity, as well as improve social and professional skills [11]. The process also encourages self-reflection and deeper understanding of the material, which may lead to improved retention and confidence [12]. Given the increasing prevalence of small group learning and a growing understanding of the benefits of peer assessments, these evaluations have become a focus of considerable amounts of research including examinations of student perceptions [13] and implementation strategies [14].

However, both students and instructors have expressed concerns about the fairness of teams and their associated peer assessments, especially due to bias [15], [16], [17], [18]. Research has shown that the experiences of women and students of color in these classrooms differ from those of their peers in terms of assessment [3]. Additionally, it is already understood that biased behaviors are commonly present in higher education classrooms [2]. Because of these concerns, interest in creating fairer peer assessments has increased. This review of the existing literature regarding bias in peer assessment is a part of a larger project undertaken to design a fairer peer assessment process. The next section of this paper will include a general discussion of bias in the classroom focused on definitions, manifestations, and measurement techniques. The following section will present evidence of bias in peer assessments and selected strategies for mitigating bias in both peer assessment and the general classroom.

TYPES OF BIAS IN THE CLASSROOM

Biased actions may result from the implicit and explicit biased attitudes of an individual. These attitudes create an individual’s subjective organizational structure for how they perceive their environment [19]. Although biased attitudes are
deeply engrained, they tend to be inconsistently expressed, depending on the social context. Explicit attitudes in particular are often moderated or “censored” in sensitive social situations [20].

In the classroom, instructors perceive incidents of implicit and explicit bias as occurring at similar frequencies [2]. Additionally, women and younger faculty members have been shown to be more likely to detect and report biased incidents [2]. Classroom bias tends to target individuals’ sexual orientation, race, sex, and ethnicity [21].

**Implicit Attitudes**

An implicit attitude is one that is “…activated by the mere presence (actual or symbolic) of the attitude object and commonly function without a person’s full awareness or control” [19, p. 62]. In this context, an “attitude object” is the target of the biased attitude. Due to this lack of awareness, implicit attitudes are inaccessible through personal introspection [22]. These attitudes shape not only the actions an individual takes, but can influence non-verbal behaviors, such as body language [19]. In short, it is difficult for an individual to recognize and address their own implicit biases without outside intervention.

In the classroom, implicit bias manifests in varying ways. One study of bias in university classrooms divided occurrences into categories of microaggressions. It was found that microassaults (exclusion), microinsults (subtle verbal snubs largely unknown to the perpetrator), and microinvalidations (negating the experiences of marginalized groups) were the most common manifestations of implicit biases in the surveyed classrooms [2]. A study by Edith Samuel [16] produced similar findings on the microaggressions experienced in the classroom. The following examples illustrate types of language and behavior that may indicate microaggressions and implicit bias:

- **Microassault:** Ignoring the contribution of a group member
- **Microinsult:** Asking a student of color “How did you get accepted here?” or “But where did you really come from?”
- **Microinvalidation:** Proclaiming to a student of color “I don’t see race” or denying any personal implicit biases

Implicit bias is also evident in unstated assumptions about other students such as needlessly offering assistance or doubting a student’s ability to complete a task [16, 2]. Additionally, manifestation of implicit bias may be observed in body language – such as an inability to make or maintain eye contact, or physically distancing oneself from a member of a disadvantaged group [23].

Unfortunately, implicit bias is not reserved for student-student interactions. This type of bias has also been seen in the marks given to students by instructors. 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) [17].

The most common method of measuring implicit bias is through the use of the Implicit Association Test (IAT). Developed in 1998 at the University of Washington, the IAT measures association of target concepts with an attribute [24]. Two concepts (such as Black vs. White) appear in a first task followed by two attributes (such as pleasant vs. unpleasant) in a second task. When the combination of concept + attribute is highly associated, participant’s categorization of stimuli items will be quicker than when the combination of concept + attribute is less associated. The difference in performance measures among the different combinations of items measures the implicit association of the concept with the attribute [24]. The IAT has also facilitated an understanding of the pervasiveness of implicit bias. Over thousands of tests, researchers have found that 80% of Whites and 40% of Blacks harbor pro-White bias [25].

**Explicit Attitudes**

An explicit attitude is one that is consciously held about a person or group. These attitudes shape responses for which individuals have the opportunity to consider the social costs and benefits of a particular action (e.g. using a homophobic slur) [19, 26]. Where implicit attitudes are difficult 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, out of the 40 students interviewed, all had experienced an incident of explicit bias. Incidences of explicit bias included: openly mocking a student’s manner of dress or religious expression (such as the dastaar or hijab), asking rude or inflammatory questions (such as “Did you live in a hut?”), and publicly ridiculing students’ abilities and accents [17]. From an instructional perspective, the incidences of explicit bias most reported by professors are the explicit use of stereotypes, telling offensive jokes, and using slurs [2].

Explicit biases are generally measured through direct self-report questionnaires. However, as biased attitudes have become less socially desirable, it has been suggested that this self-reporting cannot be assumed to be accurate, due to respondents moderating their responses to be more socially desirable (i.e. answering what they “should” say rather than what they believe). No single measurement instrument for explicit bias has emerged (in contrast with the IAT for measuring implicit bias). Due to the lack of one predominant measure and concerns about the accuracy and applicability of traditional measures [27], some researchers have created study-specific measures for assessing explicit bias (e.g. Pantos & Perkins, [28] and Devine [29]). One such study-specific measure is the “Shoulds and Woulds” scale which measures the extent to which individuals predict they would act with more prejudice than they should [29]. Nonetheless, there are a number of validated explicit bias measures in use.

One of the oldest measures of explicit bias is the Marlowe-Crowne Social Desirability Scale (MCSD). The full version of the MCSD consists of a 33-item questionnaire that
assesses beliefs and the extent to which the individual moderates these beliefs to be more socially desirable [30]. Other measures focus on attitudes toward specific groups. The widely used Modern Racism Scale (MRS) is a self-report measure of the extent of an individual’s racist beliefs [31]. This scale measures abstract ideas (i.e. affirmative action) rather than attitudes toward specific individuals or groups [32]. Similarly, the Modern Homonegativity Scale (MHS) measures contemporary attitudes toward lesbians and gay men [33]. The Attitudes Toward Women (ATW) scale measures attitudes towards the roles of women [34]. However, this scale has fallen out of favor for its dated content [35].

**BIAS IN PEER ASSESSMENT**

Bias in peer assessment can be directed at many different social and demographic groups. While the biases discussed hitherto have focused on gender, race, and ethnicity, peer assessment bias has also been observed due to social style, socioeconomic status, native language, and peer group affiliation.

Bias due to gender has been studied extensively. However, the results have been mixed for both same sex and opposite sex ratings [36], [12]. A study of oral presentation assessments found same sex bias where men rated other men slightly higher, however ratings given by women were unaffected [12]. Conversely, other researchers found that ratings between members of the same sex were consistently more prone to devaluation than when rating the opposite sex [37]. Males have been found to award the highest scores to females and females award the highest scores to males [38], [18]. Recently, though, a study of the implementation of TBL in 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 [3].

Results have again been mixed when analyzing 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 [39], [40], [41]. However, it has also been found that males receive lower ratings than their female peers [42], [43], [44]. The mixed results extend to analyses of the ratings given by male and female students. Women have, in some cases, been found to give higher evaluation scores than men [42]. Conversely, another study found men give higher evaluation scores and women give lower marks [40].

Student attitudes toward peer assessment are also inconsistent. A study of Australian undergraduates found no gender differences in satisfaction with the peer assessment process [45]. However, more recent research has shown that male students report more positive attitudes about peer assessment than female students [10], [46].

One of the few consistent findings for gender effects relates to ratings given by students to themselves. Female students consistently underestimate themselves [47], [48], [49] while male students consistently overestimate themselves [47], [48].

Many researchers have proposed explanations for these differences in results. Falchikov and Malkin [50] suggested that they may be due to gender-based communication differences and socialization or the gender association of the task or class (i.e. women in a traditionally masculine topic of study). Another explanation is the influence of ability. In many studies where women have received overall higher peer evaluation scores, they also had higher GPAs [43], [51]. It has been suggested that examining peer evaluation scores in conjunction with GPA might clarify this effect [43]. More recently, the mixed results have been suggested to be affected by cultural differences unaccounted for in analysis [18]. As the association of behaviors or personality traits with gender varies by culture, this is a plausible explanation.

Studies of racial bias in peer assessment have also returned mixed results. Multiple researchers have found that individuals tend to be rated higher by members of their own race than of other races [52], [53], [54], [55]. However, other work has demonstrated that this may not be the case. In an extensive re-analysis of Kraiger and Ford’s [55] military data, Black raters were shown to give higher ratings to White ratees than to Black ratees [56]. Further analysis of these data indicated that Black recruits consistently received lower ratings than White recruits from both Black and White raters [56]. Similarly, peer ratings in a sophomore level engineering class demonstrated that minority students received lower ratings than non-minority students [57]. Finally, in some cases, no significant evidence of racial bias has been produced [58], [59].

Recently, a study of peer assessment in large general education classes taught using TBL found convincing evidence of racial bias [3]. In these classes, students of color contributed the same number of answers and suggestions as their peers. However, they received significantly lower peer evaluation scores than White students on three out of four areas of assessment [3].

Other factors affecting the fairness of peer assessment have received less attention than gender and race. A survey of 232 undergraduates in the United States revealed that 32% would evaluate oral presentations by less wealthy students more harshly than they would evaluate more wealthy students. This effect was particularly strong for female raters. Additionally, students who held more conservative attitudes were more likely to give harsher ratings to less-wealthy students [60]. Bias due to social style has been observed in both engineering and business classrooms with students who exhibit an “expressive” social style receiving significantly higher marks than students with other social styles [61], [43]. Evidence has also suggested that peer ratings may be biased by friendship [62], [12] and language similarity (native
speakers receiving higher ratings than non-native speakers) [12]

MITIGATION

Significant research has been undertaken on bias mitigation techniques and methods. However, for the purpose of this paper, the techniques presented will be limited to those most applicable to peer assessment in a team-based classroom environment. Additionally, focus will be placed on techniques specifically intended to reduce implicit bias.

As implicit bias is inaccessible through personal introspection, it can be difficult to mitigate. Devine et al [29] posit that to begin breaking the “habit of bias”, one must first be made aware of the bias and feel concern about how the bias affects others [63], [64]. Awareness of implicit bias is generally accomplished by presenting evidence of bias via the Implicit Association Test (IAT) [65], [66].

It has been suggested that prejudice interventions yield the best outcomes when implemented in three stages: 1. Recognize the bias. 2. Apply bias mitigation techniques. 3. Allow time to practice the techniques and reflect on changes [29], [25]. From the literature, Devine et al compiled the most prominent bias mitigation strategies as a “prejudice habit-breaking intervention”. Five prominent bias mitigation techniques that have been developed are stereotype replacement, counter-stereotype imaging, individuation, perspective taking, and increasing positive contact.

Stereotype Replacement

This technique involves recognizing a stereotypical response (e.g. “women aren’t good at math”) and consciously replacing it with a rational, non-biased response (e.g. “male and female math scores show no difference when accounting for math classes taken”) [67]. Stereotype replacement involves significant self-reflection where an individual tries to determine why a stereotypical response occurred in order to determine how a biased response could be avoided in the future [29], [68].

Counter-Stereotype Imaging

Counter-Stereotype imaging involves providing an individual with a detailed description or picture of a counter-stereotype. The counter-stereotype may be abstract (e.g. “female leaders”) or specific (e.g. “Sheryl Sandberg”). While many of the techniques discussed in this section have been developed specifically in relation to implicit bias, counter-stereotype imaging has been shown to be effective at mitigating both implicit and explicit bias [69].

Individuation

Individuation prevents biased actions by learning specifics about others in order to view them as an individual as opposed to an incidence of a stereotype. Using this technique, the lines between the “in-group” and “out-group” are obscured, leading to fairer behavior [70].

Perspective Taking

Perspective taking involves imagining oneself as the target of a particular bias and contemplating the resulting psychological experience [71]. This strategy works by increasing the closeness between the individual and the targeted group [72]. Perspective taking has been shown to be very effective in reducing implicit racial bias as measured by pre- and post-intervention IAT scores [71].

Increase Positive Contact

Contact is one of the greatest influences on the strength of biases: higher levels of self-reported contact with a group result in lower implicit and explicit bias against that group [73], [74]. Positive contact is generated through interactions with the stigmatized group over a period of time. While contact over time tends to be the most effective at mitigating bias, even short periods of positive contact (such as guest speakers from targeted groups) have been shown to be effective [75].

Research on the impact of mitigation strategies on biased peer assessments in higher education is limited. However, a study that combined standard peer assessment rater training with bias reduction training produced a significant reduction in bias against individuals with differing social styles [76]. While many of these strategies have been employed in classrooms, higher education instructors often try to mitigate biased actions in other ways. A study of university professors found that few employed specific strategies and instead relied on using the bias as a discussion topic, providing a rebuttal or counterevidence, and direct confrontation [2].

CONCLUSION

Biased actions are the result of biased attitudes, which implicitly and explicitly shape the way an individual perceives their environment. In the classroom, implicit bias tends to manifest as microaggressions, negative assumptions about other students, and distancing body language, but without the awareness of the perpetrator. Explicit bias also manifests in words, interactions, and body language, but includes deliberate action and deeply held conscious beliefs. Both types of bias may affect the student’s ability to fairly assess their peer’s effort and contribution.

In peer assessments, evidence has been found to suggest that biased grades may be given based on pre-conceived notions of gender, race or ethnicity, and socioeconomic status. However, the results for both gender and racial bias are mixed and the extent to which biased peer grading occurs is still unclear.

As instructors and students have expressed a desire for fairer educational experiences, actions can and have been taken to mitigate bias in the classroom and in peer assessment. These actions typically involve bias reduction strategies and,
in the case of peer assessment, rater training combined with reduction strategies.

This review of the literature is the beginning of a larger project focused on creating fairer peer assessments by teaching students techniques to address their own biases. With this knowledge of where bias exists and the strategies used to mitigate it, the research team will develop a comprehensive peer assessment and bias mitigation methodology. This methodology will then be tested in classrooms and evaluated for its effectiveness in reducing the effects of bias on peer assessment marks. This process will focus on implementing a “bias intervention” structured to help students overcome specific biases related to race, gender, and international student status. In university settings, international students tend to be frequent targets of bias, regardless of race [16], [17]. These interventions will draw heavily from the literature on validated mitigation techniques reviewed here. After implementing the bias intervention, measures of bias as well as peer assessment scores will be recorded throughout the semester.

ACKNOWLEDGEMENTS

This material is based upon work supported by a Miller Grant from Iowa State University.

REFERENCES


