2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #50843

The Challenge of Assigning Individual Grades on Group Assignments: Incorporating Peer Assessments for Evaluating Individual Work

Lt. Col. Anthony Edward Major III, United States Military Academy
Lt. Col. William Graves, United States Military Academy

William Graves, PhD, PE is an Assistant Professor in the Department of Civil and Mechanical Engineering at the United States Military Academy, West Point, New York.

Dr. Gary A Jordan, United States Military Academy

2018 PhD Civil Engineering (Transportation Systems Engineering)- SUNY Buffalo 2017 MS Civil Engineering - SUNY Buffalo 2012 MA Economics - University of New Hampshire 2011 MBA - University of Maine 1987 BS Engineering Physics - United States Military

Major Daniel Myers, United States Military Academy

Major Daniel Myers is an instructor in the Department of Civil and Environmental Engineering at the United States Military Academy (USMA) in West Point, New York, where he teaches courses in statics, mechanics of materials, and reinforced concrete design. He graduated from USMA in 2015 with a Bachelor of Science in Civil Engineering and commissioned as an Engineer Officer. He holds a Master of Science in Civil Engineering from Virginia Tech, with a focus on structural engineering and materials. He is a licensed Professional Engineer in the state of New York.

The Challenge of Assigning Individual Grades on Group Assignments:

Incorporating Peer Assessments for Evaluating Individual Work

Abstract

By graduation, students majoring in Civil Engineering at ABET accredited institutions must demonstrate proficiency in seven student outcomes relevant to the discipline. One outcome – an ability to work effectively on a team - is particularly important in civil engineering since collaborative projects and assignments are typical course requirements. While considerable research related to the use of peer assessments within teams exists, there are gaps related to evaluating student performance using such assessments. Hence, this paper discusses our approach to evaluating individual performance on group assignments in a quantitative manner given a set of peer assessments.

Students at the United States Military Academy take a course in construction management from the institution's civil engineering program, yet students from both engineering and non-engineering disciplines enroll in the course. The diversity of academic disciplines is a result of institutional requirements for all non-engineering students to take at least three engineering courses, and many engineering disciplines outside of the civil engineering program allow their students to take the course as an elective (the course is required for those students majoring in civil engineering). Additionally, the students take the course during their junior or senior year and have already participated in group projects across their courses of study by the time they begin this course.

During the third and final block of instruction, students are divided into groups of four or five and are required to develop a construction project for life support facilities in an austere environment. At the conclusion of the project, students must qualitatively assess the performance of each member of the team on a scale ranging from excellent to poor. Given these assessments, the instructor develops a scaling factor for each individual student grade based on a ratio of the individual student's peer assessments to the group's peer assessments. At the conclusion of each semester, students were provided with an anonymous survey to collect information on their perceptions of this grading approach. This work not only describes the approach for determining the individual student grades on the project but also presents the results from the student survey.

Data was collected over the course of two non-sequential semesters with a study population of 168 students. This study will be of interest to faculty and administrators that have an interest in approaches to assigning fair grades to individual students for performance on group work

Introduction

Students graduating with a degree in Civil Engineering from ABET accredited institutions must demonstrate proficiency in seven student outcomes relevant to the discipline. One outcome – an ability to work effectively on a team - is particularly important in civil engineering. This importance stems from professional practice. Civil engineering projects, from homes to office buildings to industrial facilities and

infrastructure, often require teams with diverse talents and experiences. They also demand frequent collaboration with stakeholders who represent varied constituencies, interests, and motivations. To this end, most courses within the civil engineering curriculum at the authors' institution require group activities and assignments.

Within this curriculum, *CE450: Construction Management* at the United States Military Academy requires students to complete two significant group projects, where a considerable amount of group activity and collaboration occurs away from the classroom and beyond the eyes and ears of the teacher. Without input from student team members (i.e., peer observations and feedback), instructors have no reliable way to assess and evaluate the extent of an individual's participation and contribution outside of class. In this paper, the authors' use of peer assessment refers to student ratings and feedback of other students, whereas peer evaluation refers to the use of these assessments to quantitatively score student contributions during group projects. Many studies report variations in how feedback is used to assess or evaluate student contributions to group work [1-5].

Background

To that end, the use of student input to assess individual performance and assign individual credit is a widely researched area. Numerous studies have examined peer assessment and evaluation using a fairness-focused approach [6-12]. Most studies calculate a "score" based upon student input to measure a specific student's performance relative to other students. For example, Carson [6] directed students to rate all the group members, including themselves, on a scale of 0 to 4 for a variety of questions and then applied analysis of variance (ANOVA) to determine if there were any significant variations needing instructor scrutiny. Mathews [8] introduced a constant-sum method to score peer contribution on multiple factors related to the project (e.g., collecting data, writing the report, task and project management). In Mathews' study, each factor had a fixed number of points in which a student would distribute those points among members of the team, including themselves, based upon their evaluation of each student's contribution. Interestingly, an extensive study by Gransberg [10] in both time (14.5 years) and sample size (1,259 students) concluded that regardless of method and complexity, peer evaluations were effective in gauging and fairly assessing student contributions to group work, especially when rewarding or penalizing effort. Additionally, peer input is not the dominant factor impacting a given assignment's value or grade. When peer input is used, the weighting of such input to adjust project grades is typically minor [2-6, 13, 14].

A critical issue related to fairness is a student's ability to confidentially provide critical feedback on other students. This feedback offers additional insight to aid an instructor when assigning grades. It is important to note, however, that anonymity is typically not absolute. By design, peer ratings tend to be anonymous to students but not instructors, which helps to reinforce instructor oversight and control of the peer assessment process [1, 6, 7, 9-11, 13]. Li [15] showed that anonymity is helpful but, by itself, not necessarily sufficient in producing fairness. Li found that student concerns were measurable even when an assessment was deemed anonymous. To address these concerns, students received comprehensive training on the purpose, features, and learning benefits of assessment, along with opportunities to ask questions and engage in teacher-student and small-group discussions. With respect to overall results, Panadero and Alqassab [3] conducted a systematic review of numerous studies and reported mixed

results. They found that while anonymity increases critical feedback and may produce increased student performance, there were accuracy and bias concerns that could question assessment validity.

Abundant studies have demonstrated accuracy and bias shortcomings within peer assessments. In a study involving Information Systems students, Tu and Lu [13] used a game theory approach to find that students exaggerated the value of their personal contributions to group projects. Furthermore, they found instances of collusion in which students intentionally manipulated rankings when inappropriate incentives for completing peer assessments were applied. Klapper, Piezunka, and Dahlander found that evaluators (i.e., an individual assessing the performance of another individual) avoided offending others by biasing evaluations based upon expected reciprocity [4]. Work by Loignon et al. [16] further demonstrated that peer ratings reflected tendencies rather than true performance differences among students. Their use of the Social Relations Model showed training students with a Frame-of-Reference (FOR) approach directly addressed rater-bias concerns.

To overcome the myriads of issues related to peer assessment and evaluation, instructors have employed various strategies and methods. Elliott and Higgins [7] implemented a detailed self- and peer-assessment form that required each student to independently complete one rating sheet leveraging an itemized rating scale for every group member. The form evaluated ten specific aspects of group contribution and provided guidance on completing the form, avoiding bias, using reflective question prompts. Rather than using an itemized rating scale in which all group members could receive equally high ratings and potentially create inflated scores, Mathews [8] required students to make relative judgements by distributing a fixed amount of points for different aspects of group work. By using this approach, Mathews required students to distinguish between levels of participation, making unequal contributions visible.

Fellenz [9] developed what he terms the Groupwork Peer-Evaluation Protocol (GPEP). GPEP is a multi-method, multi-phase data collection system. It differs from other techniques by combining quantitative ratings, behavior-based checklists, evaluator (i.e., student) justifications, and reflection papers at multiple points during the group project. Because GPEP uses multiple assessment methods over time, its redundancy allows student input consistency checks and reduces the impact of bias. Other instructors have adopted a range of alternative processes – from automated grade-adjustment algorithms to long-term point allocation models – to mitigate fairness, accuracy, and rating bias concerns [10-12].

In addition to process-focused protocols and allocation methods, behaviorally anchored instruments have been validated for peer assessment on group assignments. The Comprehensive Assessment of Team Member Effectiveness (CATME) offers a theory-based framework organized around five factors – contributing to teamwork, interacting with teammates, keeping the team on track, expecting quality, and having relevant knowledge, skills, and abilities [17]. Incorporating these factors reduces ambiguity as rater judgments and interpretations are aligned with shared definitions.

To summarize, considerable research related to the use of peer assessments and evaluations within teams exists, but gaps persist related to using such assessments to evaluate performance. To address these challenges, the authors use a quantitative peer assessment process that allows each student to rate the performance of other team members (i.e., provide measurable feedback). This approach differs from other methods by using peer ratings to create an individual student's score via a scaling process. The authors contend that this process provides a degree of accountability and instructor oversight, and hence accuracy and fairness, which simply cannot be achieved without the use of peer ratings. Ultimately, this work

provides another means for the education community for appropriately awarding unique grades to individuals that are working as members of a team.

Methods

In order to develop a unique scaling factor for adjusting the group grade to an individual group member, a qualitative peer assessment form is distributed to each student (Figure 1). This peer assessment form is required to be turned in at the completion of the group project and serves as the foundation for calculating the scaling factor. Students are required to assess their own performances in addition to the performance of each individual in the group. Concurrently, the students are informed that this requirement is an assessment of contribution rather than an assessment of a grade.

		Excellent	Very Good	Satisfactory	Marginal	Deficient	No Show
Your Last Name:							
Last Name, Teammate 1:							
Last Name, Teammate 2:							
Last Name, Teammate 3:							
Last Name, Teammate 4:							
Last Name, Teammate 5:							

Figure 1. Peer Assessment Form

Students are presented with qualitative definitions (Figure 2) of the performance assessment categories ranging from *excellent* to *no show* and are informed that a rating of *satisfactory* represents a typical and honorable contribution to the group effort. While the full CATME item set is not administered in this course, category labels and definitions are consistent with its five-factor model to guide rater interpretation [18]. These individual performance assessments are not shared with the other group members to encourage the particular student to provide an honest assessment without the concern of social repercussions.

Excellent	Consistently went above and beyond, tutored group members, carried more than his/her fair share of the load
Very Good	Consistently did what (s)he was supposed to do, well prepared and cooperative
Satisfactory	Usually did what (s)he was supposed to do, acceptably prepared and cooperative
Marginal	Sometimes failed to show up or complete assignments, minimally prepared and cooperative
Deficient	Often failed to show up or complete assignments, rarely prepared
No Show	No participation at all

Figure 2. Definitions of Performance Assessment Categories

Upon collecting the peer assessment forms at the completion of the group project, the instructor determines the scaling factor for each individual student based on a predetermined point value for each performance assessment category, and a ratio of the individual student's average performance assessment to the average performance assessment of the entire group (Equation 1). This approach is intended to reduce the bias and variability in the assessments of the group members.

$$scaling\ factor = \frac{\sum_{i=1}^{n} \frac{x_i}{n}}{\sum_{j=1}^{m} \frac{x_j}{m}} \qquad \textit{(EQN 1)}$$

where

$$x \equiv point\ value\ of\ a\ given\ performance\ assessment$$

$$n \equiv total\ number\ of\ performance\ assessments\ of\ a\ given\ student$$

$$m \equiv total\ number\ of\ performance\ assessments\ of\ an\ entire\ project\ group$$

It is important to highlight that when calculating the average point value of the performance assessments for the entire student group (i.e., the denominator of *equation 1*), the performance assessments for the individual student under consideration are included.

Students are only provided with Figure 2 to guide their evaluations of fellow group members and are not provided the calculation for the scaling factor. Peer evaluations are collected immediately following the final presentation. The calculation of the scaling factor is automated using a spreadsheet, in which peer ratings are entered manually and embedded formulas compute each student's individual factor. While this process is largely manual, the authors found that the time commitment is not noticeably extended beyond the typical grading burden of a group project. Although this process could readily be transitioned to an online survey or learning management system, it is important that evaluations be completed promptly and independently to minimize opportunities for students to coordinate responses or otherwise game the system. However, it should also be noted that nothing in this approach prevents the students from "gaming the system".

Example Calculation

To illustrate this approach, assume that *Student X* submits a peer assessment form (Figure 3) for a group of 5 total students (inclusive of *Student X*).

		Excellent	Very Good	Satisfactory	Marginal	Deficient	No Show
Your Last Name:	**Student X**			✖			
Last Name, Teammate 1:	**Student A**		✖				
Last Name, Teammate 2:	**Student B**		✖				
Last Name, Teammate 3:	**Student C**	✖					
Last Name, Teammate 4:	**Student D**				✖		
Last Name, Teammate 5:							

Figure 3. Sample Peer Assessment Form

Each performance assessment is recorded by the instructor for each individual in the group, and the process is repeated for the peer evaluations collected from the remainder of the group members. In this case, the instructor had already established a scale for the performance assessments (Table 1) and collected 25 total responses with the distribution listed in Table 2.

Table 1. Point Scale for Assessments

Assessment	***Value***
Excellent	100
Very Good	90
Satisfactory	80
Marginal	70
Deficient	60
No Show	0

Table 2. Distribution of Performance Assessments for Each Student

	Excellent	***Very Good***	***Satisfactory***	***Marginal***	***Deficient***	***No Show***
Student X	1	2	2			
Student A		1	3	1		
Student B	2	2	1			
Student C		3	1	1		
Student D			1	1	3	

Given this point scale and data, the average assessment value for *Student X* is 88, while the average assessment value for the entirety of the group (inclusive of *Student X*) is 82. Following Equation 1, the scaling factor for *Student X* would be 1.07. After multiplying the scaling factor by the overall group grade on the assignment, the individual grade for *Student X* would therefore increase (as compared to *Student D*, for example, whose grade would decrease due to a scaling factor of 0.80).

Experiment

This research uses the engineering design project in *CE450: Construction Management* to implement this grading technique. In this project, students are divided into groups of five or six and are required to plan the construction of a military basecamp in an austere environment. The project occurs in the third block of instruction (out of three in total), and the students are assigned to their groups by the course instructor.

The data set for this research includes 168 students that took the course over two semesters between the fall of 2024 and 2025. The same instructor managed the course and developed grading rubrics for all assignments, but multiple instructors taught various sections of students to accommodate institutional requirements. This arrangement made it infeasible to have a single grader across the entirety of the data set.

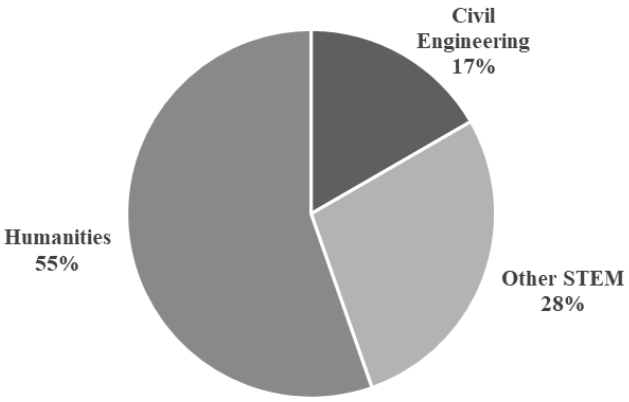

Figure 4. Distribution of Academic Majors

Another institutional requirement mandates that non-STEM academic majors enroll in at least three engineering courses. As this particular course fulfills this requirement, approximately half of the study population comes from non-STEM backgrounds (Figure 4). The civil engineering academic majors are referenced separately in Figure 4 due to the course being taught within the civil engineering program. Overall, the distribution of grades on the project (Figure 5), prior to the implementation of the scaling factor, indicates that the students generally perform well on the assignment regardless of their academic background. All students submitted a peer assessment form (Figure 1), and the point scale for the performance assessment categories follow the example from Table 1. The scores in Table 1 were selected prior to the implementation of the project to generally align with letter grades. A sensitivity analysis on the effect of these scores on the scaling factor was not conducted.

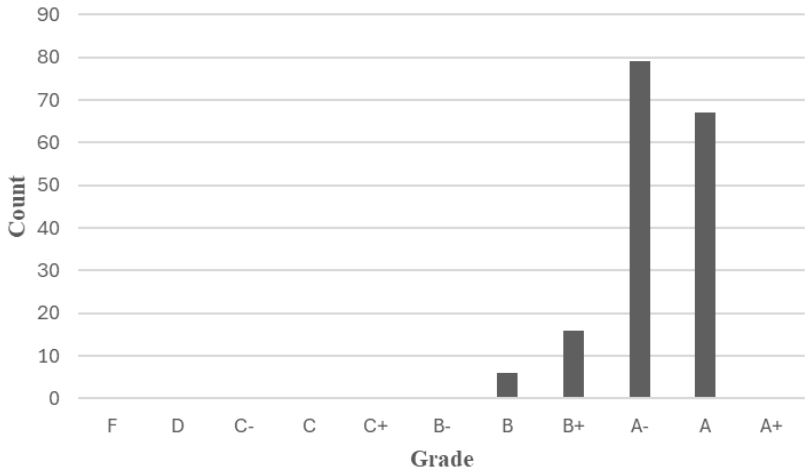

Figure 5. Distribution of Grades Prior to Implementing Scaling Factor

Following the implementation of the scaling factor, the students were allowed to answer an optional and anonymous survey asking about their perceptions of the peer assessments. The survey was intended to elicit feedback about perceived integrity of the peer assessments rather than perceived fairness of the grading approach. The following questions were presented on a five-point Likert scale with possible responses ranging from *strongly agree* to *strongly disagree*:

1. *I trust that my contributions to the Base Camp team project were fairly assessed by my peers.*
2. *My assessments of teammate contributions to the Base Camp team project were fair.*

The response rate for these questions was 84% (n = 141).

Results

The raw data for the students' grades after applying the scaling factor are shown in the following graph (Figure 6).

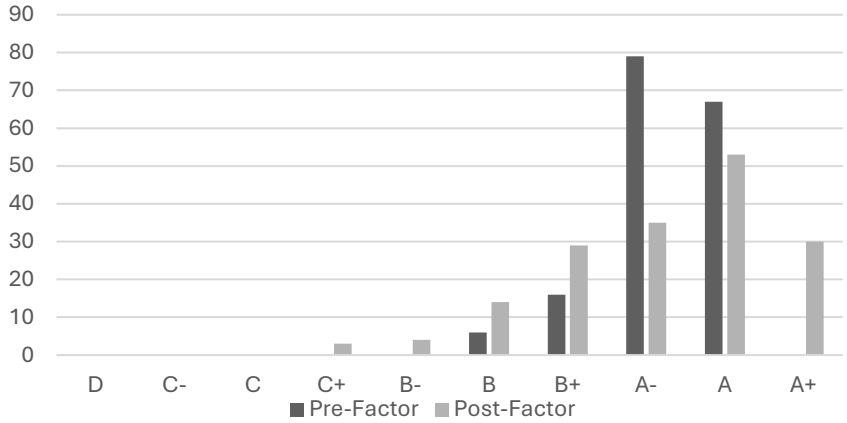

Figure 6. Distribution of Grades Prior to and After Implementing Scaling Factor

The raw data for the students' responses to the survey are shown in Figures 7 and 8.

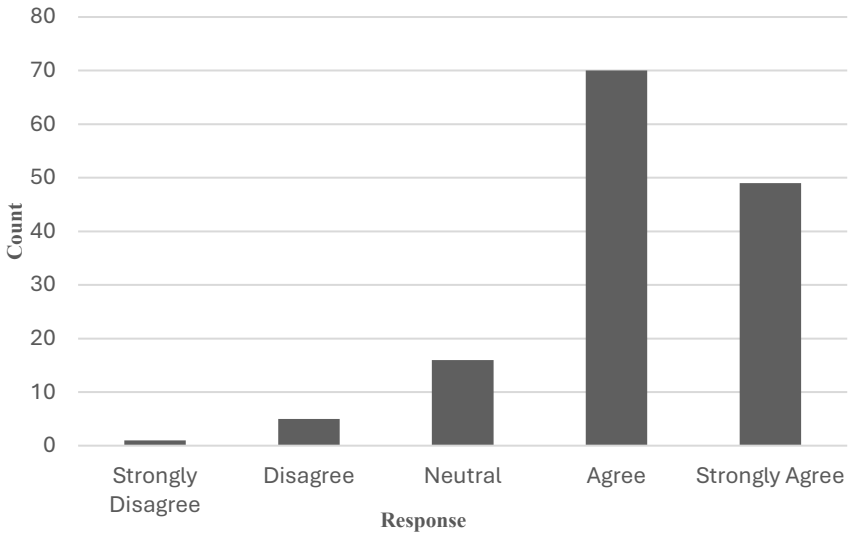

Figure 7. Distribution of Responses to the Statement: I believe my peers fairly assessed my contributions to the Base Camp team project.

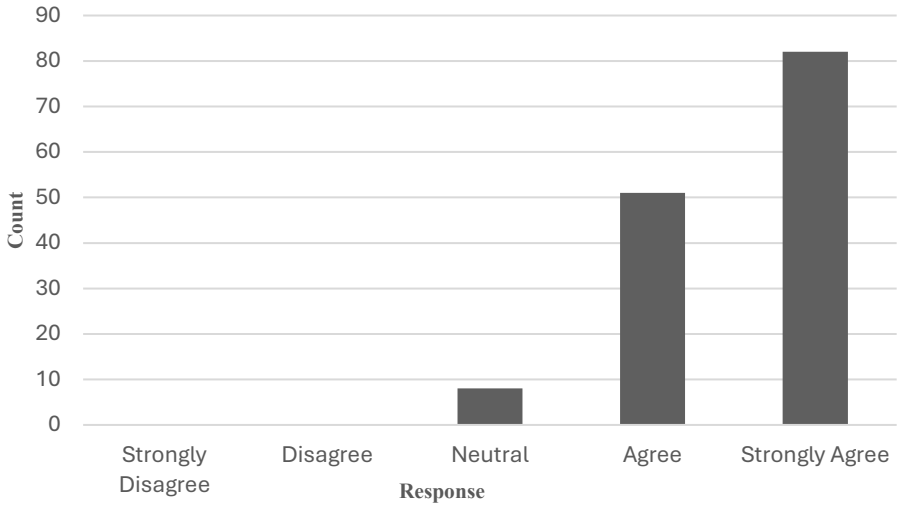

Figure 8. Distribution of Responses to the Statement: My assessments of my teammate contributions to the Base Camp team project were fair.

Discussion

The results across two non-sequential semesters indicate that incorporating a peer-based scaling factor produces a measured differentiation of individual grades within team projects while maintaining overall

performance in the grade distribution. In both semesters, post-scaling grades are concentrated in the A to B+ range with modest representation in B and B- and small counts in C+. This distribution suggests that the method adjusts individual grades without introducing excessive spread or punitive outliers. The emergence of A+ grades in each term further signals that the approach recognizes exceptional contributions when peers consistently evaluate a teammate's performance above the team norm. Quantitatively, the mean score on the assignment (92%) changes by only a fraction of a percent after implementation of the scaling factor for this given data set.

Comparing the two semesters qualitatively, the distributions after scaling are stable. Each semester shows similar proportions of A, A-, and B+ outcomes with only minor counts in lower categories. This consistency implies that the presented approach is robust to cohort composition and instructional section differences, even when multiple instructors grade various sections of the course.

Student perceptions of the peer assessment process are an important complement to the grade outcomes. On the statement "I trust that my contributions to the Base Camp team project were fairly assessed by my peers," responses are overwhelmingly positive across both semesters and across Civil Engineering and non–Civil Engineering majors, with most students selecting Agree or Strongly Agree and very few selecting Disagree or Strongly Disagree. Similarly, for "My assessments of teammate contributions were fair," the pattern remains favorable, though a small number of students selected Neutral, which warrants attention to rater training and clarity of the rating anchors. Overall, the perception data suggests strong acceptance of the process, with isolated pockets of skepticism that can potentially be mitigated with clearer instructions. Also, these responses appear to support the point that students largely provided honest feedback without attempting to manipulate the process to improve final grades.

Conclusion and Future Work

Teamwork is central to Civil Engineering curricula, yet translating group performance into individual grades remains an unresolved challenge. This study standardizes peer assessments into a clear scaling mechanism that adjusts individual grades relative to team norms, providing faculty with a practical tool to recognize contribution without abandoning collective outcomes.

- Exceptional contributors are recognized, evidenced by the appearance of A+ outcomes, when peers consistently rate performance above the team's mean.
- Underperformers experienced downward adjustments, but these changes were moderate and did not result in extreme penalties, suggesting the scaling factor effectively differentiates contributions without disproportionately harming lower-performing students.
- Student acceptance is strong. Most respondents across majors agree or strongly agree that peer ratings were fair, for both their own contributions and their ratings of teammates.
- Implementation choices, such as clear descriptions of each choice when rating their peers and the inclusion of the graded student in the group average, likely contributed to consistency and bounded adjustments.

Several variations can be implemented to adapt the method to different instructional contexts. For example, instructors may impose grade caps and floors, such as limiting upward adjustments so that no

student's grade exceeds 100% or restricting changes to within one letter grade of the original group score. Downward adjustments can also be bounded to prevent disproportionate penalties. Aggregation methods can be refined by using medians or trimmed means instead of simple averages, or by dropping extreme ratings to reduce bias. Weighting schemes may normalize ratings by team size or assign greater weight to raters who demonstrate consistency. Categories within the peer assessment can be behaviorally anchored and weighted according to importance, such as collaboration, timeliness, and deliverable quality. Procedural safeguards, such as requiring rationales for extreme ratings, can further enhance fairness. Finally, transparency measures such as sharing aggregate team feedback without revealing individual ratings can promote reflective practice while maintaining anonymity. These options allow instructors to tailor the approach while preserving its core principle of reasonable recognition of individual contributions.

Future work should examine the breakdown of academic majors within the course to determine whether disciplinary background influences rating patterns or perceptions of fairness. Expanding the qualitative survey to include more focused questions—such as why students felt their assessments were fair or unfair—would provide richer insight into attitudes and potential biases. Additional studies could explore how modifying the differential between peer feedback scores affects the impact of evaluations on final grades, as this was not investigated as part of this research effort. These investigations would not only improve the robustness of the grading method but also enhance its adaptability across diverse instructional contexts.

References

[1] J. Alden, "Assessment of Individual Student Performance in Online Team Projects," *Online Learning,* vol. 15, no. 3, 2011.

[2] D. A. Dickerson, S. Masta, M. W. Ohland and A. L. Pawley, "Is Carla Grumpy? Analysis of Peer Evaluations to Explore Microaggressions and Other Marginalizing Behaviors in Engineering Student Teams," *Journal of Engineering Education,* no. 113, pp. 603-634, 2024.

[3] E. Panadero and M. Alqassab, "An Empirical Review of Anonymity Effects in Peer Assessment, Peer Feedback, Peer Review, Peer Evaluation and Peer Grading," *Assessment & Evaluation in Higher Education,* vol. 44, no. 8, pp. 1253-1278, 2019.

[4] H. Klapper, H. Piezunka and L. Dahlander, "Peer Evaluations: Evaluating and Being Evaluated," *Organization Science,* vol. 35, no. 4, pp. 1363-1387, 2024.

[5] E. Morris and D. A. McAdams, "Integration and Evaluation of Peer Grading in a Graduate-level Engineering Design Course," in *American Society of Engineering Education Annual Conference*, Columbus, 2017.

[6] J. Carson, "Evaluation Of Individual Performance On Group Projects," in *2008 Annual Conference & Exposition*, Pittsburgh, PA, 2008.

[7] N. Elliott and A. Higgins, "Self and Peer Assessment – Does It Make A Difference To Student Group Work?," *Nurse Education in Practice,* vol. 5, no. 1, pp. 40-48, 2005.

[8] B. P. Mathews, "Assessing individual contributions: Experience of peer evaluation in major group projects," *British Journal of Educational Technology,* vol. 25, no. 1, pp. 19-28, 1994.

[9] M. R. Fellenz, "Toward Fairness in Assessing Student Groupwork: A Protocol for Peer Evaluation of Individual Contributions," *Journal of Management Education,* vol. 30, no. 4, pp. 570-591, 2006.

[10] D. D. Gransberg, "Quantifying the Impact of Peer Evaluations on Student Team Project Grading," *International Journal of Construction Education and Research,* vol. 6, no. 1, pp. 3-17, 01 March 2010.

[11] N. Holland and L. Feigenbaum, "Using Peer Evaluations to Assign Grades on Group Projects," *Journal of Construction Education,* vol. 3, no. 3, pp. 182-188, 1998.

[12] A. M. Vollstedt and E. L. Wang, "A Method for Adjusting Group-Based Grades," in *2011 Pacific Southwest Section Meeting*, Fresno, CA, 2011.

[13] Y. Tu and M. Lu, "Peer-and-Self Assessment to Reveal the Ranking of Each Individual's Contribution to a Group Project," *Journal of Information Systems Education,* vol. 16, no. 2, pp. 197-206, 2005.

[14] M. L. Wagner, D. C. Suh and S. Cruz, "Peer- and Self-Grading Compared to Faculty Grading," *American Journal of Pharmaceutical Education,* vol. 75, no. 7, pp. 1-7, 2011.

[15] L. Li, "The Role Of Anonymity In Peer Assessment," *Assessment & Evaluation in Higher Education,* vol. 42, no. 4, pp. 645-656, 2017.

[16] A. C. Loignon, D. J. Woehr, J. S. Thomas, M. L. Loughry, M. W. Ohland and D. M. Ferguson, "Facilitating Peer Evaluation in Team Contexts: The Impact of Frame-of-Reference Rater Training," *Academy of Management Learning & Education,* vol. 16, no. 4, pp. 562-578, 2017.

[17] M. L. Loughry, M. W. Ohland and D. D. Moore, "Development of a Theory-Based Assessment of Team Member Effectiveness," *Educational and Psychological Measurement,* vol. 67, no. 3, pp. 505-524, 2007.

[18] M. W. Ohland, M. L. Loughry, D. J. Woehr, L. G. Bullard, R. M. Felder, C. J. Finelli, R. A. Layton, H. R. Pomeranz and D. G. Schmucker, "The Comprehensive Assessment of Team Member Effectiveness: Development of a Behaviorally Anchored Rating Scale for Self- and Peer Evaluatin," *Academy of Management Learning & Education,* vol. 11, no. 4, pp. 609-630, 2012.