WIP From Notebooks to Contribution Tables: Addressing Fairness and Accountability in Student Team Projects

Dr. Jorge Ivan Rodriguez-Devora, University of Georgia

Dr. Rodriguez serves as the industry capstone project coordinator for the College of Engineering at the University of Georgia. He is a faculty member of the School of Environmental, Civil, Agricultural and Mechanical Engineering.

Dr. Roger Hilten, University of Georgia

Roger Hilten is an Assistant Professor of Practice at the University of Georgia in the College of Engineering's School of Environmental, Civil, Agricultural, and Mechanical Engineering. Dr. HIlten is deeply involved in Capstone Design at all levels, from project solicitation to individual student mentoring. Dr. Hilten collaborates with a team of instructors to develop and deliver Capstone course material while gathering data to continually improve the Capstone experience for students, project sponsors, and faculty involved. Administratively, Dr. Hilten works with the Capstone Industry Coordinator to seek out new project sponsors and develop projects to fit the needs of Capstone while also supporting students' on-the-ground procurement, prototyping, and testing efforts for over 100 Capstone projects. Dr. Hilten acts the Director for the College's Lab Support Team which manages instructional, fabrication, and testing facilities as well as the tools and equipment that support students to effectively complete all types of projects, Capstone Design projects included.

Dr. Wayne Johnson, University of Georgia

Wayne M. Johnson is a Senior Lecturer in the School of Environmental, Civil, Agricultural and Mechanical Engineering at the University of Georgia (UGA) in Athens, GA. Prior to joining UGA in 2022, he was a Professor of Mechanical Engineering at Georgia Southern University-Armstrong Campus, Savannah GA. He received his Ph.D. and M.S. in Mechanical Engineering from Georgia Institute of Technology and his B.S. in Mechanical Engineering (Cum Laude) from Louisiana State University. He has published 16 papers in peer-reviewed journals, 28 papers in peer-reviewed conference proceedings, and given 12 technical presentations on various topics including: additive manufacturing, mechatronics, biomechanics, and engineering education. He currently teaches the Engineered Systems In Society, Mechanical Engineering Professional Practice, and Capstone Design I and II courses.

Kevin J. Wu, The University of Georgia

Kevin J. Wu is currently the Engineering Catalyst for Innovation Gateway at the University of Georgia and an Adjunct Instructor with the University of Georgia College of Engineering. Mr. Wu's current research interests lie in examining the factors that affect innovation in undergraduate teams. Mr. Wu received a bachelor's and master's degree in engineering from the University of Georgia.

WIP: From Notebooks to Contribution Tables: Addressing Fairness and Accountability in Student Team Projects

Abstract

Engineering education increasingly emphasizes teamwork, yet ethical challenges arise when evaluating individual contributions within group projects. Traditional reliance on engineering notebooks and other design documentation as assessment tools often introduces ambiguity regarding authorship, originality, and equitable credit assignment. This study examines the ethical tensions that emerge when instructors evaluate student notebooks to infer individual effort, particularly the risks of misrepresentation, overgeneralization, or unequal recognition.

To investigate these issues, a mixed methodological approach combining analytical and qualitative methods was employed. The analytical component focused on systematically examining contribution tables and notebook entries across multiple project milestones to quantify patterns of participation, authorship, and deliverable ownership. The qualitative component involved thematic coding of instructor reflections, student interviews, and peer feedback to capture perceptions of fairness, transparency, and ethical awareness in team assessment. This dual approach enabled a nuanced understanding of both the measurable distribution of effort, and the interpretive processes instructors use when assessing individual work within collaborative contexts.

Instructor analysis played a central role in this study. Faculty reviewers were trained to apply standardized rubrics for evaluating the completeness, coherence, and authorship clarity in student notebooks, while also documenting their own ethical reasoning and uncertainties during the grading process. These reflective analyses provided insights into how evaluators navigate the balance between accountability and trust, and how structured documentation tools may reduce subjective bias in assessing teamwork.

Practical implications of the anticipated findings include the potential for broader adoption of contribution tracking frameworks that support both formative feedback and ethical grading practices in design-based courses. Such frameworks can help institutions promote fairness, transparency, and ethical conduct while reinforcing professional competencies aligned with ABET outcomes on teamwork and responsibility.

This study is original in its integration of ethical analysis, instructor reflection, and structured data collection to address the persistent challenge of individual assessment in collaborative engineering design. By bridging pedagogical ethics and design documentation practices, it offers an evidence-based model for cultivating integrity and accountability within engineering education.

Tags:

Ethics, teamwork, design documentation, accountability, contribution tracking, engineering journaling

Introduction

Engineering education increasingly relies on team-based project work to develop students' technical competence alongside professional skills such as collaboration, communication, and shared responsibility. However, the shift toward teamwork creates persistent ethical challenges for instructors who must evaluate individual performance within collective deliverables. The central tension is that teamwork requires interdependence, yet grading practices often demand individual accountability. This challenge is especially pronounced in design-based courses, where assessment is frequently tied to documentation artifacts, most notably engineering notebooks, design logs, and milestone reports, that are used to infer individual effort and ownership.

Engineering notebooks serve an important professional function by supporting traceability, decision justification, and responsible recordkeeping. Training models for notebook management emphasize ethical expectations such as accurate documentation, transparent revision practices, and appropriate attribution, framing notebook habits as part of responsible engineering conduct rather than merely a grading requirement [1]. Yet in educational settings, notebooks can become ethically ambiguous instruments for assessment: they may capture only partial evidence of labor, disproportionately reward students who document more rather than contribute more, or unintentionally penalize students whose work is less visible in written artifacts. Early team evaluation approaches in design education recognized the importance of notebooks as evidence of participation but also highlighted the difficulty of using documentation alone to evaluate multidisciplinary team contributions [2].

To address inequities in evaluating contribution, engineering education has increasingly adopted peer evaluation systems designed to individualize teamwork assessment. Instruments such as CATME provide structured mechanisms for capturing team member effectiveness and distributing credit more equitably than group grades alone [3]. Related work demonstrates that behavioral anchors can improve the reliability of peer ratings, strengthening the defensibility of contribution judgments and reducing arbitrary interpretation [4]. Recent approaches extend these methods through continuous peer evaluation and rubric-based scoring frameworks that aim to improve transparency, reduce free-riding, and provide actionable feedback across project milestones [5-6]. Nevertheless, peer evaluation also introduces ethical risks, including bias, interpersonal retaliation, and distorted perceptions of competence, especially when ratings become high-stakes determinants of individual grades [7].

These concerns intersect with broader findings in engineering academic integrity research, which underscore that students' ethical decision-making is shaped by assessment pressures, collaboration norms, and ambiguous boundaries around authorship and acceptable teamwork practices [8]. Taken together, prior work suggests that fair assessment in team projects requires more than "better grading"—it requires systems that clarify authorship, reduce subjectivity, and support ethical accountability.

Team-based design courses are essential for preparing engineering students for professional practice, yet they introduce persistent ethical and pedagogical challenges in fairly evaluating individual contributions. When assessment relies heavily on team deliverables, unequal participation may go undetected, weakening accountability and undermining academic integrity. Professional engineering practice depends on traceable documentation and clear attribution of work; therefore, educational assessment methods must reflect these same ethical standards. Although prior research has demonstrated the value of peer evaluation and structured assessment tools, fewer studies have examined how engineering notebooks, contribution tables, and instructor interpretation interact to influence fairness in grading decisions. Addressing this gap is necessary to ensure that assessment systems accurately recognize individual effort, support equitable grading, and reinforce professional responsibility. By examining documentation practices alongside peer and instructor evaluations, this study provides evidence to inform more transparent, reliable, and ethically grounded approaches to assessing individual contributions within collaborative engineering design courses.

Methodology

This study examined documentation-based and evaluation-based indicators of individual contribution within the two-semester senior Capstone Design sequence in AENG and MCHE at XXX. In this course series, student teams collaborate with a capstone partner to complete a sponsored project. Data were derived from course participants in the 2024–2025 academic year (AY24–25), including 6 AENG students and 198 MCHE students distributed across 44 projects. This study was reviewed by the Institutional Review Board and determined to be exempt under federal regulations (45 CFR §46.104(d)(1)) as research conducted in established educational settings involving normal instructional practices, with all data de-identified prior to analysis.

Data were gathered across both the Fall and Spring semesters of AY24–25. The dataset included quantitative course artifacts and performance measures, along with qualitative feedback and instructor commentary. Quantitative data sources included, standardized contribution tables included in final reports (Fall and Spring), engineering notebooks and logged hours (Fall and Spring), instructor/supervisor grades for: individual contribution (Fall and Spring) and notebook performance, 360° evaluation scores (1–5 scale), including: Relative Peer Performance (RPP) and peer average score. RPP captured the relationship between peer evaluation and self-assessment. For example, if a student rated themselves as 4 and peers rated them as 4, the RPP equals 1. Qualitative data sources included, open-ended 360° peer comments (Fall and Spring), instructor/supervisor comments drawn from course records (e.g., project health notes, self-evaluation feedback, discussion boards, or related course communications)

Team selection for focused analysis. A subset of teams was selected for deeper analysis based on peer-evaluation patterns. First, teams were screened for the presence of at least one student with RPP < 0.8, which was treated as a potential indicator of disagreement between self-perception and peer perception of contribution. In total, 23 teams contained at least one student below this threshold. However, only three teams included two or more students with RPP < 0.8; these three teams were selected for targeted analysis due to the concentration of low-RPP values within the same team. In addition, three teams were selected that contained substantial peer-review commentary and exhibited RPP values near 1, reflecting a close alignment between self-assessment and peer ratings. Team selection details are summarized in Table 1.

Table 1. Teams included in the study.

Team ID	Team	Team average peer score (n=44, 4.21 +/- 0.47) Individual min/max = 3/5	Entries under threshold (0.8)	Team average relative peer performance (RPP)	Team size (members)
	Amicable Teams				
1	F24 – Industry project on manufacturing	4.54	4	1	5 (1-5)
2	F24 – Process optimization industry project	4.99	3	1	3 (6-8)
3	F24 – Industry Plant drawing update project	4.205	6	1	4 (9-12)
	Teams at Risk				
4	F24 – Competition project	3.43	3	0.71	4 (13-16)
5	F24 – Industry project on Construction field	3.81	3	0.72	5 (17-21)
6	F24 - University Veterinary project	3.56	4	0.695	5 (22-26)

Notebooks were evaluated at the midterm and final of each semester using a standardized 5-point rubric. Notebook grading was conducted by three instructors. The rubric included three criteria:

1. Evidence of progress (1 point): progress demonstrated through notebook entries and hours logged relative to teammates and course expectations
2. Organization (1 point): clarity, formatting, and legibility
3. Evidence of ABET Outcome 1 (3 points): documentation of engineering/scientific tools and methods applied during the project (e.g., decision matrices, QFD, interviews, calculations, simulations, sketches, ideas, and material selection)

Logged hours verification. Students reported project hours throughout the semester. These hours were reviewed and verified by course graders to support consistency and reduce reporting errors.

Quantitative analysis relied on group comparisons across teams and selected team subsets. Measures included contribution table patterns, notebook grades, logged hours, instructor/supervisor contribution grades, and 360° evaluation metrics (RPP and peer average score). Comparisons were used to examine whether teams exhibiting low-RPP clustering or high-RPP divergence also differed on documentation-based and instructor-assessed indicators of contribution. Qualitative analysis examined peer comments and instructor/supervisor reflections to characterize perceived fairness, transparency, and contribution disputes within teams. Comments associated with selected teams were reviewed to identify recurring themes related to accountability, documentation clarity, and interpretation of individual effort within collaborative work.

Results and Discussion

Team-based projects develop professional collaboration skills, but they also create opportunities for unequal labor distribution and ambiguous authorship, particularly when grading emphasizes team deliverables. Consistent with prior engineering education work on team assessment and accountability [3], [5], this study analyzed multiple evidence streams—contribution tables, notebooks, logged hours, instructor ratings, and 360° evaluations—to identify patterns associated with fairness concerns and team conflict.

Across the six teams examined (Figure 1; black boxes denote team membership), logged hours and notebook entries were the strongest predictors of contribution conflict risk. The three "at-risk" teams exhibited clearer mismatches between documented effort (hours/notebook evidence) and team expectations than the three amicable teams. This finding aligns with the role of engineering notebooks as traceable records of work progress and professional accountability [1] but also reinforces that notebooks alone can be incomplete evidence if they are inconsistently maintained or used primarily as a compliance artifact.

The 360° evaluation provided the clearest early signal of ethical issues related to contribution. When applying a threshold score of 3 out of 5, the 360° evaluation captured 100% of the underperforming learners identified through other evidence streams. This supports prior work showing that structured peer evaluation tools can strengthen fairness and individual accountability in team grading [3].

However, qualitative patterns differed sharply by team type and semester. Amicable teams provided minimal peer comments, and when present, comments were consistently affirmative and aligned with positive team dynamics. In contrast, at-risk teams generated substantial first-semester commentary, particularly from team leaders, describing concerns related to effort, reliability, and engagement. Despite this strong first-semester signal, a significant reduction in commentary occurred in the second semester, suggesting evaluation fatigue or reduced willingness to document team conflict. This decline limits the reliability of end-of-year peer feedback as a standalone mechanism for accountability, consistent with concerns that peer evaluation quality can degrade when participation drops [5].

Contribution tables strengthened accountability by making reported work visible to all team members, not only instructors. Because entries are shared within the team, disengagement can be detected as milestones progress—particularly when sections remain vague, incomplete, or unclaimed. Teams with stronger documentation practices (more hours logged and more notebook evidence) also tended to produce more detailed contribution tables, suggesting that consistent documentation supports clearer attribution and reduces ambiguity about ownership.

At the same time, several cases demonstrated that low contribution could persist despite weak documentation. Students with limited logged hours, minimal notebook entries, and poor contribution table reporting could still pass when team deliverables earned high grades. This reflects a recurring integrity challenge in engineering education: when team outputs dominate grading, individual accountability becomes structurally weak [8]. A key ethical concern emerged when students with minimal hours logged, sparse notebook evidence, and weak contribution table reporting still received passing outcomes due to the heavy weighting of team-based milestone grades. Figure 2 shows that grade reductions did not consistently reflect low participation. For example, in the case of students 22 and 23, contribution tables in the final report indicated little to no documented ownership of report sections, and cells were left blank or unsigned. This reflects a broader academic integrity challenge: accountability mechanisms can fail when individual evidence is weak but team outputs remain sufficient for credit [8].

Instructor reflective analysis highlighted a grading dilemma: some instructors were more generous with highly weighted team-based grades, which reduced the impact of smaller individual components (e.g., notebook, contribution, and 360° scores). This practice inflated overall grades for teams with uneven participation and diminished fairness compared with teams where instructors weighted individual evidence more strictly. These inconsistencies can unintentionally reinforce free-riding incentives and weaken ethical expectations for authorship and responsibility.

Implications: earlier intervention through midterm peer evaluation.

Given the strong detection performance of peer evaluation (threshold ≤3/5) and the observed drop in second-semester commentary, the findings support implementing a midterm 360° evaluation each semester. A midterm checkpoint would enable timely intervention when students fall below 3 out of 5, including structured coaching, clarification of expectations, and documentation requirements before team outcomes are finalized. This approach aligns with continuous peer evaluation models intended to support formative correction rather than end-stage punishment [5], and it strengthens the fairness rationale behind peer evaluation systems used in engineering teams [3].

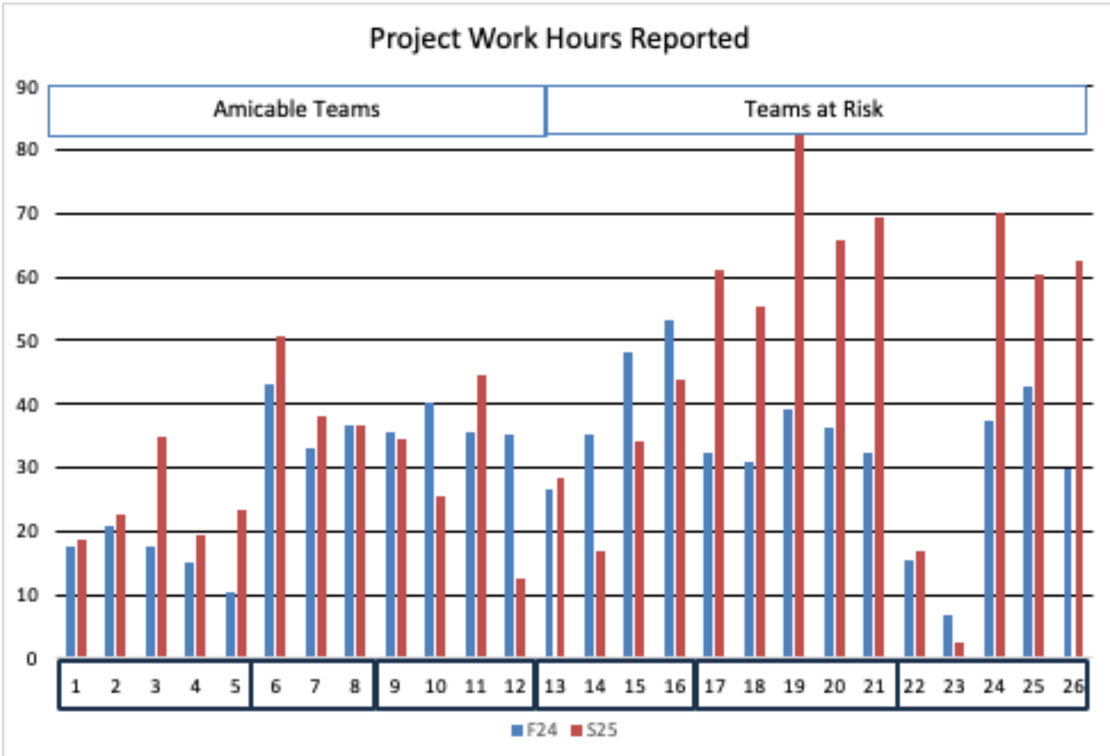

Figure 1. Self-reported project hours. Black boxes indicate team membership. Teams on left were identified as amicable and teams on right as "at risk" of contribution conflict risk.

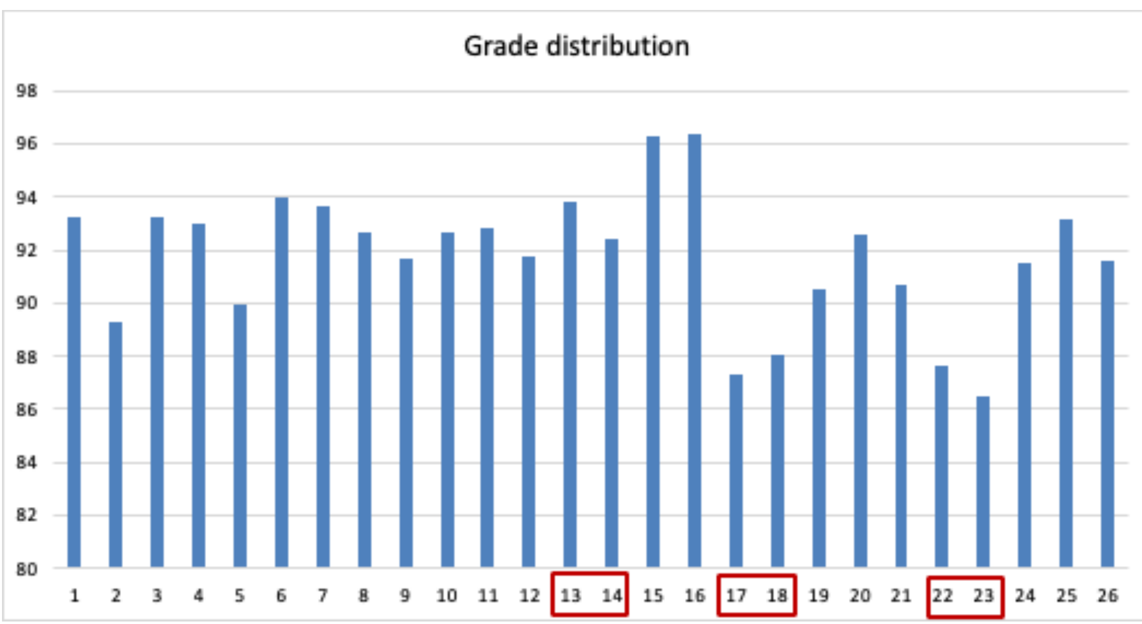

Figure 2. Final grade distribution across team members (out of 100). Learners identify on red boxes were flag both by at least 5 indicators, including individual contribution, notebooks, logged hours, and 360- evaluation (fall/spring).

In future work, we will implement and evaluate structured interventions to improve the fairness and accountability of how we assess individual contributions in team-based design courses. As a key next step, we will introduce mid-semester 360° evaluations as formative checkpoints to identify students whose contribution scores are at or below 3 out of 5, enabling instructors to intervene in a timely manner with targeted feedback and documentation review. In parallel, we will refine the notebook structure and contribution tables—including standardized entry classifications, explicit attribution requirements, and verified team sign-offs—and pilot these changes to strengthen traceability and clarify authorship. We will conduct a longitudinal comparison between pre- and post-intervention cohorts to examine alignment among peer evaluation scores, documented hours, notebook assessments, contribution reporting, and instructor-assigned grades, and we will pair this analysis with qualitative data on instructor and student perceptions of fairness. Through these efforts, we aim to develop and validate an evidence-based accountability framework that integrates documentation and peer evaluation to support more transparent, consistent, and ethically grounded assessment of individual contributions in collaborative engineering design.

Acknowledgments

The authors thank the Capstone Design program and the Engineering Education XXX Institute (CCC) at the College of Engineering at the University of XXXX.

References

[1] S. Hills and C. J. Light, "Training for responsible and ethical management of lab notebooks in a course-based undergraduate research experience," *Journal of Microbiology & Biology Education*, vol. 23, no. 2, Art. no. e00024-22, 2022, doi: 10.1128/jmbe.00024-22.

[2] L. Slivovsky, W. Oakes, and L. Jamieson, "Evaluating multidisciplinary design teams," in *Proc. 2003 ASEE Annual Conference*, Nashville, TN, USA, Jun. 2003, doi: 10.18260/1-2--11942.

[3] M. W. Ohland, H. R. Pomeranz, and H. W. Feinstein, "The comprehensive assessment of team member effectiveness: A new peer evaluation instrument," in *Proc. 2006 ASEE Annual Conference & Exposition*, Chicago, IL, USA, Jun. 2006, doi: 10.18260/1-2--751.

[4] M. W. Ohland, R. A. Layton, M. L. Loughry, and A. G. Yuhasz, "Effects of behavioral anchors on peer evaluation reliability," *Journal of Engineering Education*, vol. 94, no. 3, pp. 319–326, 2005, doi: 10.1002/j.2168-9830.2005.tb00856.x.

[5] W. A. Friess and A. J. Goupee, "Using continuous peer evaluation in team-based engineering capstone projects: A case study," *IEEE Transactions on Education*, vol. 63, no. 2, pp. 82–87, May 2020, doi: 10.1109/TE.2020.2970549.

[6] S. Thite, J. Ravishankar, I. Tomeo-Reyes, and A. M. Ortiz, "Design of a simple rubric to peer-evaluate the teamwork skills of engineering students," *European Journal of Engineering Education*, vol. 49, no. 4, pp. 623–646, 2024, doi: 10.1080/03043797.2024.2338239.

[7] 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*, vol. 113, no. 3, pp. 603–634, 2024, doi: 10.1002/jee.20606.

[8] C. J. Finelli, T. S. Harding, D. D. Carpenter, and M. J. Mayhew, “Academic integrity among engineering undergraduates: Seven years of research by the E^3 Team,” in *Proc. 2007 ASEE Annual Conference & Exposition*, Honolulu, HI, USA, Jun. 2007.