How Engineering Students Describe Good Teammates: A Thematic Analysis of Peer Evaluations Across Multiple Capstone Cohorts

Dr. Tracy Anne Hammond, Texas A&M University

Dr. Tracy Anne Hammond is a Professor of Computer Science and Engineering at Texas A&M University and Director of the Sketch Recognition Lab. She earned her Ph.D. in Electrical Engineering and Computer Science from the MIT Computer Science and Artificial Intelligence Laboratory at Massachusetts Institute of Technology, along with additional degrees from Columbia University in computer science, mathematics, applied physics, and anthropology.

Her research focuses on artificial intelligence, sketch recognition, intelligent user interfaces, haptics, and engineering education, with applications in education, health, accessibility, and related domains. She contributes to research, teaching, and professional service in AI and engineering education, with an emphasis on responsible and inclusive design.

Jobin Varughese, Texas A&M University

Jobin Varughese is a doctoral student in the College of Engineering at Texas A&M University, where he also serves as Senior Educational Data Analyst at the Center for Teaching Excellence. He holds an M.S. in Management Information Systems with an emphasis in Data Science from Oklahoma State University and a B.S. in Electronics and Communication Engineering from Christ University. His research draws on learning analytics, AI in education, and human-computer interaction to observe, build, and study systems that learn how people learn. In this work, he partners closely with faculty, instructional staff, and students — believing that the most meaningful scholarship emerges not from research alone, but from collaborations with the people closest to the learning experience.

How Engineering Students Describe Good Teammates: A Thematic Analysis of Peer Evaluations Across Multiple Capstone Cohorts

Abstract

This full research paper analyzes 2,489 peer and self-evaluation responses from senior engineering capstone design courses across 14 evaluation waves and four cohort years (2019–2025) to examine how students evaluate teammate effectiveness in their own words. Using an inductive thematic analysis of open-ended peer and self-evaluation comments alongside descriptive summaries of productivity ratings, we identify eight recurring themes in student-authored evaluations of teammate performance. Contrary to assumptions common in engineering education research, students most frequently describe effective teammates in terms of concrete technical outputs rather than communication or collaboration process. Peer and self-evaluations diverge systematically, with students describing their own professional conduct far more prominently than peers do. Thematic structure was stable across all cohort years including the recent emergence of generative AI tools, with saturation reached early and sustained through 2025. These findings provide a student-centered taxonomy of teammate behaviors, reveal gaps between student evaluative language and professional teamwork frameworks, and suggest ways to refine peer evaluation instruments and instructional practices to surface under-recognized competencies.

Introduction

What does a student actually notice when evaluating a teammate? Not what they are asked to notice, not what the rubric rewards, but what surfaces naturally when given an open prompt and a real project on the line. This question is worth taking seriously. Peer evaluation is among the most widely used mechanisms for developing teamwork competencies in engineering education, yet the evaluative language students produce in authentic settings has received comparatively little systematic attention. Teamwork is a foundational outcome in engineering education, emphasized in accreditation criteria, industry expectations, and project-based curricula [1, 2]. Capstone courses rely on team-based work to approximate professional engineering practice and to assess students' ability to collaborate, communicate, and contribute to shared goals [3, 4]. Prior research has examined the reliability and validity of peer ratings and their alignment with instructor assessments [5]. However, the qualitative content of peer feedback—particularly across multiple cohorts and repeated evaluation cycles—remains underexplored. Analyzing student-authored peer comments offers a unique opportunity to surface how students recognize and value behaviors associated with effective teammates and the norms that shape team dynamics over time.

This study does not claim to measure teamwork holistically; rather, it examines how students represent and evaluate teammate performance through peer evaluation discourse. We analyze authentic peer and self-evaluations from senior engineering capstone courses across four cohort years (2019–2025), collected as part of routine course practice rather than elicited for research purposes.

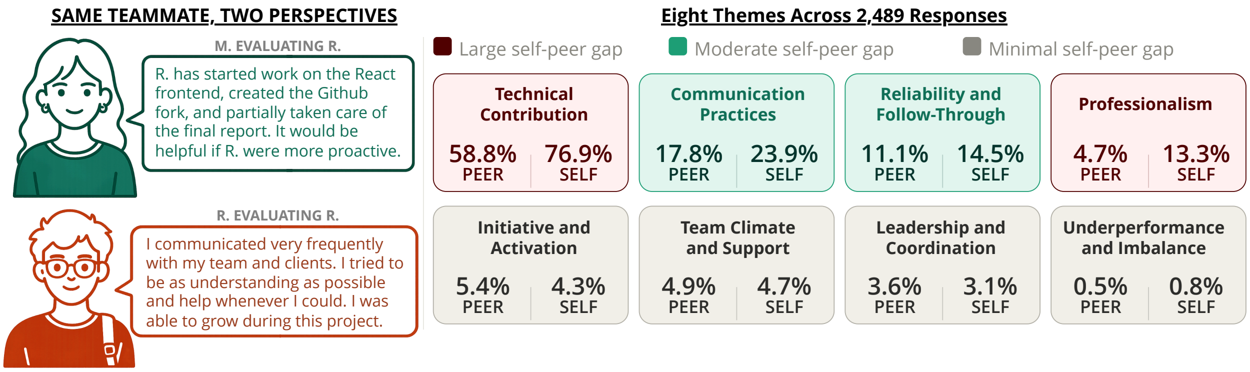

FIGURE 1. Peer and self-evaluations of the same teammate illustrate systematic differences in evaluative language. While peer evaluations emphasize observable technical outputs, self-evaluations foreground relational effort and professional identity.

We adopt an inductive qualitative approach to examine how students evaluate teammate effectiveness in their own words. This work makes three contributions. First, it provides a student-centered taxonomy of teammate behaviors grounded in authentic peer discourse from capstone design teams, revealing that technical contribution dominates student evaluative language in ways that existing peer assessment frameworks do not anticipate. Second, it highlights alignments and gaps between student-authored evaluations and competencies commonly emphasized in professional and educational frameworks (see Figure 1). Third, it offers actionable implications for capstone instruction, including refinements to peer evaluation prompts and strategies for surfacing teamwork competencies that students may not naturally articulate.

This study addresses three research questions:

(RQ1) What themes characterize how engineering students evaluate teammate effectiveness in peer evaluations?

(RQ2) How do these themes differ between peer- and self-descriptions?

(RQ3) How do thematic patterns relate descriptively to peer-assigned performance ratings?

Related Work

Teamwork is a core learning outcome in engineering education, emphasized in accreditation criteria and embedded in capstone design courses that center authentic, team-based problem solving [1, 4, 6]. Prior work has drawn from industrial and organizational psychology to conceptualize team effectiveness, including models of team development [7], collective performance [8], frameworks of high-performing teams [9], and psychological safety as a key enabling condition for team learning [10]. Tannenbaum and Salas offer an integrative framework organizing team effectiveness around seven drivers—communication, coordination, cooperation, capabilities, conditions, coaching, and cognitions—providing a multidimensional vocabulary for examining what makes teams function [11]. In engineering education, these frameworks have been adapted to examine student team processes and outcomes, focusing on observable behaviors such as communication, coordination, leadership, and task execution [2, 9].

Several assessment instruments have been developed to evaluate teamwork in student teams. CATME provides structured, behaviorally anchored peer evaluation scales aligned with teamwork constructs from organizational psychology [12]. Other studies examine peer assessment practices more broadly, highlighting both their reliability and their role in promoting learning through reflection and feedback [13–15]. Research in engineering contexts emphasizes peer evaluation as a mechanism for accountability, formative feedback, and professional skill development [4–6, 12]. CliftonStrengths is a strengths-based self-assessment tool developed by Gallup that identifies individuals' dominant talent themes across 34 categories [16]; it is widely used in engineering and organizational contexts, including in the capstone courses examined here.

Much of this literature operationalizes teamwork through predefined categories or rating scales derived from expert models. While these approaches support consistency and comparability, they reflect researcher-defined constructs applied to student behavior rather than student-generated constructs derived from authentic evaluative language. Prior qualitative work in engineering education has explored student experiences of teamwork, communication, and peer feedback processes [14, 15], but tends to treat teamwork as one component of broader learning processes rather than as a focal construct grounded in student language. Research on learning and feedback further suggests that the meaning students make of evaluative practices is central to their educational impact, particularly when students articulate judgments in their own words rather than through numerical scores alone [14, 17, 18].

Prior research has also linked individual traits such as conscientiousness to peer ratings in team settings [19, 20], and actor-observer asymmetries in behavior explanation offer a theoretical lens for understanding why self- and peer-evaluations systematically differ [21, 22]. Actors explaining their own behavior tend to emphasize reasons and intentions, while observers rely on visible actions and outcomes. This asymmetry, well established in social psychology, has not been applied systematically to engineering peer evaluation data across multiple cohorts. Organizational justice research further suggests that fairness and equity concerns become salient in evaluative discourse primarily when norms are violated [23], which may help explain why themes related to workload imbalance appear infrequently in routine peer feedback.

Relatively little work has systematically analyzed the open-ended language students use in peer evaluations across multiple teams and time points. Existing studies often prioritize psychometric validation or alignment with instructor-defined constructs [12, 13], leaving student-authored qualitative feedback underexplored as a data source in its own right. This study addresses that by conducting an inductive thematic analysis of open-ended peer evaluation comments across multiple semesters and evaluation waves, foregrounding the language students produce rather than the constructs researchers impose.

Context and Data Collection

Data were collected from senior-level engineering capstone design courses at a large public research university in the United States. These courses center on team-based, industry- or stakeholder-sponsored projects completed within a single academic semester. Teams of four to six students manage the full project lifecycle, including requirements gathering, implementation, testing, and sponsor communication. Teamwork is an explicit learning objective, and peer evaluation is used formatively throughout the semester to support reflection, accountability, and team development.

Peer evaluation prompts included open-ended questions asking students to describe teammates' strengths, communication practices, and contributions. Although wording and ordering varied slightly across course offerings and academic years, the core intent remained consistent: to elicit student-authored evaluations of teammate behaviors grounded in ongoing project work. Earlier cohorts (2019) used a rubric-style form with numeric ratings and limited open-ended fields; later cohorts (2023–2025) used a form with dedicated open-ended prompts for communication, contributions, strengths-based reflection (e.g., CliftonStrengths), and general feedback. Because this study focuses on patterns in how students evaluate teammates rather than on instrument-specific differences, all qualitative responses were analyzed together rather than segmented by form version.

All available peer evaluation responses from spring semester offerings between 2019 and 2025 were merged into a unified dataset. Each response included metadata (academic year, semester, evaluation wave, timestamp), and duplicate entries were removed prior to analysis. The resulting corpus includes 2,489 responses across 14 evaluation waves spanning four cohort years, including both peer evaluations ($n = 2{,}234$) and self-evaluations ($n = 255$), enabling analysis of stable patterns in student evaluative language rather than one-time reflections. See Table 5 for the full distribution by year and wave.

Peer evaluation data were collected as part of routine instructional practice and were not originally intended for research. Prior to analysis, all responses were anonymized and identifying information removed. This study was reviewed and approved by the Institutional Review Board under protocol STUDY2026-0095, consistent with institutional guidelines for the secondary use of educational records.

Methods

This study employs an inductive qualitative design to analyze open-ended peer evaluation comments, with descriptive quantitative summaries used to contextualize observed patterns [24]. The goal is not to test predefined constructs but to characterize patterns in how students articulate evaluative judgments within authentic educational contexts.

Coding followed a multi-cycle inductive process consistent with qualitative coding practices described by Saldaña [25], including initial open coding followed by pattern coding to consolidate themes across the dataset. Trustworthiness was addressed using criteria established by Lincoln and Guba [26], including credibility through iterative engagement with the dataset, dependability through consistent analytic procedures and audit trail documentation, and transferability through thick description of the institutional context and corpus characteristics. In accordance with Creswell and Poth [24], analytic validity was supported through sustained engagement with the data across multiple coding cycles and transparent documentation of code development and theme boundary decisions.

The unit of analysis was an idea unit within each peer evaluation response. Responses were conceptually segmented to allow multiple themes to be identified within a single comment, enabling comparison across responses of varying length and specificity. Codes were grounded in observable behaviors and practices rather than inferred traits. When students used

TABLE 1. Theme Frequencies Across Peer and Self-Evaluations

Theme	Peer N	Peer %	Self N	Self %	Mean prod.
Technical Contribution	1,313	58.8	196	76.9	4.67
Communication Practices	397	17.8	61	23.9	4.49
Reliability and Follow-Through	249	11.1	37	14.5	4.63
Initiative and Activation	121	5.4	11	4.3	4.49
Team Climate and Support	110	4.9	12	4.7	4.60
Professionalism	105	4.7	34	13.3	4.65
Leadership and Coordination	80	3.6	8	3.1	4.74
Underperformance and Imbalance	11	0.5	2	0.8	3.44

Note. *Peer $N = 2{,}234$; Self $N = 255$. Responses could reflect multiple themes; percentages do not sum to 100%. Mean productivity ratings are peer-assigned (1–5 scale). Underperformance co-occurs with the lowest mean productivity rating (3.44) and the lowest rate of high ratings (44.4%), consistent with its role as a marker of perceived contribution failure.*

CliftonStrengths labels [16], these were not treated as analytic constructs; instead, accompanying behavioral descriptions were coded. Codes were iteratively refined and grouped into broader thematic categories through axial consolidation. Themes were retained based on recurrence across cohorts, conceptual clarity, and analytic salience. Thematic saturation was assessed by tracking the emergence of new themes across evaluation waves; all eight themes had emerged by the first wave of the 2023 cohort, and no new themes appeared across the subsequent ten waves through 2025. To address RQ3, descriptive statistics were computed for numeric peer ratings and theme co-occurrence; these analyses are descriptive only and not intended for inferential testing.

Results

Across cohorts and evaluation waves, students consistently evaluated teammates using a shared set of behavioral and relational constructs. Eight themes emerged from the inductive analysis (Tables 1 and 2): Technical Contribution, Communication Practices, Reliability and Follow-Through, Initiative and Project Activation, Team Climate and Interpersonal Support, Professionalism, Leadership and Coordination, and Underperformance and Imbalance. Themes were stable across cohorts and project types, with thematic saturation reached by the first wave of the 2023 cohort and sustained through 2025.

RQ1: Themes in Student Peer Evaluations

Technical Contribution was the most prevalent theme, appearing in 58.8% of peer evaluations and co-occurring with a mean productivity rating of 4.67. Students overwhelmingly evaluated teammates in terms of concrete, named project outputs: code written, hardware built, models trained, documentation produced, and deliverables completed. This dominance was consistent across all cohort years and grew across the semester, from 53.1% in early waves to 63.8% in later waves, suggesting that as project demands intensified, students focused on who delivered tangible work. Technical contribution was rarely discussed in isolation; students emphasized whether teammates communicated their progress and integrated their work with others.

TABLE 2. Summary of Themes and Exemplar Quotes by Evaluation Wave

Theme	Early-wave quote (waves 1–2)	Late-wave quote (waves 3–5)
Technical Contribution	"F. has put many many hours in working with T. to implement data collection, which is a tough task."	"C. worked on frontend, testing different models, restyling pages, and ensuring the project went along smoothly."
Communication Practices	"He always lets us know what he is working on and responds quickly in our Discord."	"J. is incredibly communicative and vocal about his progress and struggles."
Reliability and Follow-Through	"M. is accountable. He always finishes his work on time and leaves no room for error."	"It does feel like it might have been procrastinated a bit until the last minute."
Initiative and Activation	"He has shown Activator as he often takes initiative in starting tasks."	"When J. is given a task, he does it without question until it's finished. He even took the initiative to go further."
Leadership and Coordination	"As project manager, she has started developing the vision for our team and been consistent in developing our required workloads."	"A. really owned the role as project manager in the last few weeks. She set clear needs and deadlines."
Team Climate and Support	"He did a great job at ensuring that everyone was included in the group discussions."	"Working with him as a team was incredibly difficult; it felt like the design had to go exactly as he said."
Professionalism	"H. has been a great communicator with team members and the client."	"Communication by leading the interactions between our team and industry sponsor."
Underperformance and Imbalance	"R. does seem like he wants to help, he just hasn't done much."	"A. hasn't done anything outright for the project. I would recommend he find something to do and start working on it."

Note. *All quotes reproduced verbatim from the corpus; lightly edited only for anonymization where required. Names replaced with first initial only to protect student privacy. Early = evaluation waves 1–2; Late = evaluation waves 3–5.*

Communication Practices was the second most prevalent theme (17.8%). Students emphasized responsiveness, frequency of updates, and proactive sharing of progress and blockers rather than talkativeness or eloquence. Communication was framed as maintaining shared awareness within the team. High productivity ratings frequently co-occurred with descriptions of timely updates, while lower ratings referenced delayed responses or lack of visibility into a teammate's work.

Reliability and Follow-Through appeared in 11.1% of peer evaluations. High-performing teammates were described as individuals who completed tasks independently, met deadlines without reminders, and could be counted on. In contrast, lower-rated evaluations emphasized last-minute work, missed meetings, and uneven contribution.

Initiative and Activation (5.4%) captured whether teammates proactively sought and started work without direction, often referenced through CliftonStrengths labels such as Activator or Achiever [16] accompanied by behavioral descriptions. CliftonStrengths labels referenced most frequently by students included Analytical, Responsibility, and Restorative rather than Communication, suggesting that students recognized problem-solving and accountable behaviors as prominent teammate strengths even when given an explicit prompt to name a strength (see Appendix B, Table 6).

Leadership and Coordination (3.6%) appeared in both formal project manager roles and informal coordination behaviors such as organizing meetings, assigning tasks, and keeping the team on schedule. Leadership was framed pragmatically rather than hierarchically; students emphasized coordination work over positional authority, consistent with distributed leadership perspectives [2, 9, 11].

Team Climate and Support (4.9%) captured specific interpersonal behaviors shaping the team environment, including helping teammates through problems, maintaining morale under pressure, and managing conflict. Vague positive statements were excluded from this theme; only responses describing specific climate-relevant behaviors were coded. *Professionalism* (4.7%) referenced conduct in external-facing contexts, primarily sponsor and client communication. *Underperformance and Imbalance* was the rarest theme (0.5%), co-occurring with the lowest mean productivity rating (3.44) and the lowest rate of high ratings (44.4%). When present, it reflected explicit concern about contribution adequacy or workload equity, and appeared primarily in later evaluation waves.

RQ2: Peer vs. Self-Descriptions

Comparisons between self-evaluations ($n = 255$) and peer evaluations ($n = 2{,}234$) revealed systematic differences across all themes (Table 1; Figure 1). Self-descriptions showed higher rates across nearly every theme, with the sharpest divergence in Professionalism (13.3% self vs. 4.7% peer, a gap of 8.6 percentage points). Students described their own professional conduct in external settings far more prominently than their peers described theirs, suggesting that students perceive their own professional contributions as more visible than observers do. Technical Contribution also showed a substantial self-peer gap (76.9% vs. 58.8%), while Team Climate and Leadership showed minimal divergence.

These patterns align with actor-observer asymmetries in behavioral explanation [21, 22]: actors explaining their own behavior tend to emphasize reasons and internal motivations—goals, effort, intentions—while observers rely on enabling factors such as visible actions and outputs. Self-evaluations foreground intention and professional self-perception; peer evaluations foreground output and reliability. Together, peer and self-assessments provide a more complete picture of teammate performance than either alone.

RQ3: Thematic Patterns and Performance Ratings

Descriptive analysis of peer-assigned productivity ratings revealed consistent patterns across themes. Technical Contribution and Leadership co-occurred with the highest mean productivity ratings (4.67 and 4.74 respectively), while Underperformance co-occurred with the lowest (3.44). Communication and Initiative showed similar mean ratings (4.49), suggesting that students associate both responsiveness and proactivity with comparable levels of perceived productivity. Meaningful variance in ratings emerged over time: later evaluation waves exhibited greater dispersion and more critical commentary, consistent with qualitative shifts toward diagnostic feedback as project demands intensified. Full descriptive statistics are reported in Appendix B. Theme co-occurrence with productivity ratings is summarized in Table 1.

Discussion

The dominant finding is that Technical Contribution accounts for the majority of evaluative language across all waves and years. When given open-ended prompts, students default to describing who built what. Communication, reliability, and initiative follow at substantially lower rates. This pattern holds even as project demands shift across the semester: Technical Contribution grows from 53.1% in early waves to 63.8% in later waves, while communication-related language decreases slightly. Students increasingly attend to deliverables as project stakes intensify. This challenges a common assumption in engineering education that peer evaluation naturally surfaces process-oriented teamwork competencies. It does not. Left to their own language, students evaluate output.

Mapping student themes against Tannenbaum and Salas's [11] seven-driver framework makes visible what students notice and what remains outside their evaluative view (Table 3). In the table, green cells indicate that a student theme strongly addresses that driver; yellow cells indicate partial alignment; blank cells indicate absence. Student-generated themes address communication, coordination, cooperation, and capabilities. Conditions and coaching appear only partially, while cognitions are entirely absent from student evaluative language across all 2,489 responses. No student described whether the team had adequate structural support, whether teammates were developing each other's skills, or whether the team held a shared mental model of the project. The red Cognitions column makes this visible at a glance: no student theme addresses shared mental models, transactive memory, or awareness of who knows what on the team. It reflects what peer evaluation prompts and student evaluative frames currently make visible.

The divergence between peer and self-evaluations adds a further dimension. Self-evaluations showed higher rates across nearly all themes, with the sharpest gap in Professionalism (13.3% self vs. 4.7% peer). Students perceive their own professional conduct in external settings as more prominent than peers do. This aligns with actor-observer asymmetries in behavioral explanation [21, 22]: actors emphasize intentions and internal effort; observers emphasize visible outputs and reliability. Together, peer and self-assessments provide a more complete picture of teammate performance than either alone.

One finding warrants particular attention for its implications beyond this dataset. Thematic saturation was reached by the first wave of the 2023 cohort and sustained through every subsequent wave including 2025. The evaluative norms students apply to teammates did not change across a period that included the mainstreaming of generative AI into undergraduate work. Students in 2025 using AI-assisted development still evaluated teammates on the same dimensions as students in 2019. What students notice about collaboration appears more stable than the tools that mediate it.

Implications for Practice

The findings point to several concrete opportunities for instructors and program designers working with capstone peer evaluation.

The dominance of Technical Contribution in student evaluative language suggests that peer evaluation instruments are currently optimized for surfacing what students already notice. Prompts that ask students to describe contributions naturally elicit output-focused responses.

TABLE 3. Student Themes Mapped to Tannenbaum and Salas's [11] Seven Cs of Team Effectiveness

Theme	Comm.	Coord.	Coop.	Capab.	Cond.	Coach.	Cogn.
Technical Contribution							
Communication Practices							
Reliability and Follow-Through							
Leadership and Coordination							
Initiative and Activation							
Team Climate and Support							
Professionalism							
Underperformance and Imbalance							

Strongly present Partially present Absent Not surfaced by students

Note. *Cognitions (Cogn.) was the only driver entirely absent from student evaluative language across all 2,489 responses and four cohort years (2019–2025). Conditions and Coaching appeared only partially, through Team Climate and Leadership respectively. No student described whether the team held a shared mental model of the project, suggesting that shared cognition lies entirely outside the evaluative frame that current peer evaluation prompts activate. Comm. = Communication; Coord. = Coordination; Coop. = Cooperation; Capab. = Capabilities; Cond. = Conditions; Coach. = Coaching; Cogn. = Cognitions.*

Instruments should more explicitly prompt attention to less visible but professionally critical competencies. Targeted questions asking how teammates handle disagreement, support equitable participation, or contribute to shared team understanding can broaden evaluative frames and legitimize these behaviors as valued aspects of effective collaboration [11, 12].

The peer-self divergence in Professionalism suggests that students perceive their own external-facing conduct more prominently than peers do. This asymmetry, consistent with actor-observer differences in behavior explanation [21, 22], has practical implications for how instructors interpret self-assessments. Self-evaluations should not be read as independent corroboration of peer ratings; they reflect a systematically different evaluative frame.

Peer evaluation comments represent a rich instructional artifact beyond their role in accountability. Instructors can use anonymized excerpts to surface shared norms, illustrate effective behaviors, and prompt discussion about what professional teamwork frameworks identify as important but students do not spontaneously name [2, 24]. The absent and underrepresented drivers in Table 3 are a practical starting point: structured reflection activities built around Cognitions in particular, and more deliberate attention to Conditions and Coaching, can help students develop evaluative vocabulary for dimensions of team effectiveness they currently cannot see.

Because students tend to equate effectiveness with visible output and responsiveness, quieter or less visible contributions risk being systematically undervalued in peer feedback. Instructors and mentors should interpret peer feedback with this bias in mind, particularly when making grading or intervention decisions [13, 14].

Limitations

This study has several limitations. Data were drawn from capstone courses at a single institution; although the corpus spans multiple cohorts and evaluation waves, results may reflect institution-specific norms, instructional practices, or disciplinary expectations, and replication across institutions would strengthen transferability [26]. The peer evaluation instruments included strengths-oriented prompts referencing CliftonStrengths [16], which may have shaped how students framed feedback, particularly by encouraging strengths-based or positively valenced descriptions, and may have contributed to the relative scarcity of conflict or equity language. The 2019 cohort used a rubric-style form with more limited open-ended fields than the 2023–2025 cohorts, introducing structural variation that may affect comparability across years.

Conclusion

This study examined how engineering students evaluate teammate effectiveness through authentic peer evaluations collected across four cohort years and 14 evaluation waves in capstone design courses. Analysis of student-authored feedback revealed a consistent and perhaps counterintuitive pattern: students overwhelmingly evaluate who delivered tangible work. Technical contribution dominated evaluative language at more than three times the rate of any other theme. Communication, reliability, and initiative followed at substantially lower rates. Competencies central to professional teamwork frameworks, including conflict management, coaching, shared cognition, and structural conditions, were largely absent from student language regardless of cohort year, project type, or semester timing.

The peer-self divergence adds a further dimension. Students describe their own professional conduct and technical contributions more prominently than peers do, consistent with actor-observer asymmetries in behavioral explanation [21, 22]. Together, peer and self-assessments capture complementary aspects of teammate performance that neither alone provides.

Perhaps the most consequential finding for the field is the temporal stability of these patterns. Thematic saturation was reached by the first wave of the 2023 cohort and sustained through every subsequent wave including 2025. The evaluative norms students apply to teammates did not change across a period that included the mainstreaming of generative AI tools into undergraduate engineering work. What students notice about collaboration appears more durable than the technologies that mediate it. This stability suggests that the taxonomy identified here reflects something fundamental about how students recognize and assess effective teamwork, and that surfacing the competencies students do not naturally articulate will require deliberate instructional intervention rather than assuming that evaluative frames evolve on their own.

References

[1] ABET, “Criteria for accrediting engineering programs, 2022–2023,” ABET, Tech. Rep., 2023, available: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2022-2023/.

[2] M. Borrego, J. Karlin, L. D. McNair, and K. Beddoes, “Team effectiveness theory from industrial and organizational psychology applied to engineering student project teams: A research review,” *Journal of Engineering Education*, vol. 102, no. 4, pp. 472–512, 2013, doi: 10.1002/jee.20023.

[3] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, “Engineering design thinking, teaching, and learning,” *Journal of Engineering Education*, vol. 94, no. 1, pp. 103–120, 2005, doi: 10.1002/j.2168-9830.2005.tb00832.x.

[4] J. J. Pembridge and M. C. Paretti, “The current state of capstone design pedagogy,” in *Proceedings of the 2010 ASEE Annual Conference & Exposition*. Louisville, KY: ASEE Conferences, 2010, doi: 10.18260/1-2–16048.

[5] C. M. Cestone, R. E. Levine, and D. R. Lane, “Peer assessment and evaluation in team-based learning,” *New Directions for Teaching and Learning*, vol. 2008, no. 116, pp. 69–78, 2008, doi: 10.1002/tl.334.

[6] S. Howe and J. Wilbarger, “2005 national survey of engineering capstone design courses,” in *Proceedings of the 2006 ASEE Annual Conference & Exposition*. Chicago, IL: ASEE Conferences, 2006, doi: 10.18260/1-2–236.

[7] B. W. Tuckman, “Developmental sequence in small groups,” *Psychological Bulletin*, vol. 63, no. 6, pp. 384–399, 1965, doi: 10.1037/h0022100.

[8] J. R. Katzenbach and D. K. Smith, *The Wisdom of Teams*. Boston, MA: Harvard Business School Press, 1993.

[9] E. Salas, D. E. Sims, and C. S. Burke, “Is there a “big five” in teamwork?” *Small Group Research*, vol. 36, no. 5, pp. 555–599, 2005, doi: 10.1177/1046496405277134.

[10] A. Edmondson, “Psychological safety and learning behavior in work teams,” *Administrative Science Quarterly*, vol. 44, no. 2, pp. 350–383, 1999, doi: 10.2307/2666999.

[11] S. I. Tannenbaum and E. Salas, *Teams That Work: The Seven Drivers of Team Effectiveness*. New York, NY: Oxford University Press, 2021, doi: 10.1093/oso/9780190056964.001.0001.

[12] 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 evaluation,” *Academy of Management Learning & Education*, vol. 11, no. 4, pp. 609–630, 2012, doi: 10.5465/amle.2010.0177.

[13] N. Falchikov and J. Goldfinch, “Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks,” *Review of Educational Research*, vol. 70, no. 3, pp. 287–322, 2000, doi: 10.3102/00346543070003287.

[14] N.-F. Liu and D. Carless, “Peer feedback: The learning element of peer assessment,” *Teaching in Higher Education*, vol. 11, no. 3, pp. 279–290, 2006, doi: 10.1080/13562510600680582.

[15] I. van den Berg, W. Admiraal, and A. Pilot, “Peer assessment in university teaching: Evaluating seven course designs,” *Assessment & Evaluation in Higher Education*, vol. 31, no. 1, pp. 19–36, 2006, doi: 10.1080/02602930500262346.

[16] T. Rath, *StrengthsFinder 2.0*. New York, NY: Gallup Press, 2007.

[17] D. J. Nicol and D. Macfarlane-Dick, “Formative assessment and self-regulated learning: A model and seven principles of good feedback practice,” *Studies in Higher Education*, vol. 31, no. 2, pp. 199–218, 2006, doi: 10.1080/03075070600572090.

[18] D. Nicol, A. Thomson, and C. Breslin, “Rethinking feedback practices in higher education: A peer review perspective,” *Assessment & Evaluation in Higher Education*, vol. 39, no. 1, pp. 102–122, 2014, doi: 10.1080/02602938.2013.795518.

[19] C. C. Martin and K. D. Locke, “What do peer evaluations represent? A study of rater consensus and target personality,” *Frontiers in Education*, vol. 7, p. 746457, 2022, doi: 10.3389/feduc.2022.746457.

[20] M. Ohme and H. Zacher, “Job performance ratings: The relative importance of mental ability, conscientiousness, and career adaptability,” *Journal of Vocational Behavior*, vol. 87, pp. 161–170, 2015, doi: 10.1016/j.jvb.2015.01.001.

[21] B. F. Malle, “The actor–observer asymmetry in attribution: A (surprising) meta-analysis,” *Psychological Bulletin*, vol. 132, no. 6, pp. 895–919, 2006, doi: 10.1037/0033-2909.132.6.895.

[22] B. F. Malle, J. Knobe, and S. E. Nelson, “Actor–observer asymmetries in explanations of behavior: New answers to an old question,” *Journal of Personality and Social Psychology*, vol. 93, no. 4, pp. 491–514, 2007, doi: 10.1037/0022-3514.93.4.491.

[23] J. A. Colquitt, “On the dimensionality of organizational justice: A construct validation of a measure,” *Journal of Applied Psychology*, vol. 86, no. 3, pp. 386–400, 2001, doi: 10.1037/0021-9010.86.3.386.

[24] J. W. Creswell and C. N. Poth, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, 4th ed. Thousand Oaks, CA: Sage, 2018.

[25] J. Saldaña, *The Coding Manual for Qualitative Researchers*, 3rd ed. Thousand Oaks, CA: Sage, 2016.

[26] Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*. Beverly Hills, CA: Sage, 1985.

Appendix A: Analytic Memo and Codebook

A.1 Analytic Purpose and Coding Process

This memo documents the procedures used to develop themes and select exemplar quotes. Coding followed an inductive, multi-cycle process consistent with Saldaña [25]. Initial coding involved open coding to identify recurring concepts in student-authored language related to teammate behaviors, contributions, and evaluative judgments. Subsequent cycles involved pattern coding to group related codes into higher-level categories. Constant comparison was used to refine code definitions, resolve overlap, and ensure conceptual clarity. Analytic memoing was conducted throughout to document coding decisions and emerging interpretations. An audit trail consistent with Lincoln and Guba [26] was maintained across all coding cycles.

A.2 Analytic Boundaries and Trustworthiness

This analysis is descriptive and interpretive. Trustworthiness was supported through iterative coding across multiple cycles, analytic memoing, transparent documentation of theme development and boundary decisions, and maintenance of an audit trail consistent with Lincoln and Guba [26]. Themes were retained based on recurrence across cohorts, conceptual clarity, and analytic salience. Saturation was assessed by tracking theme emergence across evaluation waves; all eight themes had emerged by the first wave of the 2023 cohort, and no new themes appeared across the subsequent ten waves through 2025.

A.3 Full Codebook with Theme Definitions and Subcodes

TABLE 4. Theme Definitions and Subcodes

Theme	**Definition**	**Subcodes**
Technical Contribution	Student describes specific technical or project outputs: code written, hardware built, models trained, named artifacts, tools learned, or concrete deliverables. The operative distinction is work-named vs. work-unnamed, not code vs. non-code.	Named software or hardware task; specific deliverable produced; tool or framework learned for project; integration of work with team; volume or significance of contribution described.
Communication Practices	Student evaluates how the teammate communicates: responsiveness, frequency of updates, clarity, platform use, or proactive sharing of progress and blockers. Includes negative instances.	Responds promptly to messages; keeps team updated on progress; communicates clearly or proactively; platform use (Discord, GroupMe, etc.); fails to communicate or goes quiet.
Reliability and Follow-Through	Student evaluates whether the teammate completes commitments, meets deadlines, shows up consistently, and can be counted on without reminders. Includes negative instances.	Completes assigned tasks; meets deadlines or is on time; consistent and dependable; works independently without reminders; misses meetings or submits last-minute work.
Initiative and Activation	Student describes teammate proactively taking on work, moving the project forward without direction, or going beyond assigned tasks. Includes negative instances.	Takes initiative without being asked; self-starts or activates project momentum; goes above and beyond; waits to be assigned rather than seeking work.
Leadership and Coordination	Student describes teammate organizing, coordinating, or managing how the team works: formal PM roles or informal coordination behaviors.	Formal project manager role; organizes or runs meetings; assigns or delegates tasks; keeps team on track or on schedule; coordinates between team members.
Team Climate and Support	Student describes specific behaviors shaping the team environment: helping teammates, mediating conflict, maintaining morale, or including others. Vague positive statements excluded.	Helps or supports specific teammates; positive attitude visibly affecting morale; conflict management or mediation; including isolated members; disruptive or hostile behavior.
Professionalism	Student evaluates professional conduct in external-facing contexts: sponsor or client communication, representing the team, professional etiquette in formal settings.	Professional communication with sponsors or clients; represents team well externally; professional conduct in formal settings; unprofessional conduct noted.
Underperformance and Imbalance	Student expresses concern about inadequate contribution, unequal workload distribution, or behavioral disruptions affecting team equity.	Not contributing enough; workload inequity or unfairness; one person carrying too much; disruptive or hostile behavior.

Note. *Theme definitions and subcodes were developed inductively through Saldña's [25] multi-cycle coding process. A response could be coded for multiple themes. Vague positive statements (e.g., "great teammate") were excluded from all themes and tracked separately as non-specific praise.*

Appendix B: Supporting Data

TABLE 5. Sample Size by Year, Semester, and Evaluation Wave

Year	Semester	Wave	N
2019	Spring	1	144
2019	Spring	2	133
2019	Spring	3	234
2023	Spring	1	160
2023	Spring	2	160
2023	Spring	3	150
2023	Spring	4	153
2023	Spring	5	159
2024	Spring	1	168
2024	Spring	2	138
2024	Spring	3	148
2024	Spring	4	147
2025	Spring	2	331
2025	Spring	3	309
Total			**2,489**

Note. *Peer evaluations $n = 2{,}234$; self-evaluations $n = 255$. The 2019 cohort used a rubric-style form with more limited open-ended fields than the 2023–2025 cohorts. The 2025 cohort includes waves 2 and 3 only; wave 1 data were not available for that year.*

TABLE 6. Most Frequently Referenced CliftonStrengths Labels (Peer and Self-Evaluations Combined)

Strength label	Mentions
Analytical	130
Responsibility	86
Restorative	82
Achiever	75
Consistency	69
Adaptability	63
Learner	55
Focus	55
Communication	54
Harmony	53

Note. *Counts reflect all peer and self-evaluations combined ($N = 2{,}489$). Analytical ranked first, not Communication, contrary to the assumption that students most frequently invoke communication-related strengths when describing effective teammates. CliftonStrengths labels described by Rath [16].*