Evaluating Assessment Practices in Team-Based Computing Capstone Projects

Sara Hooshangi*
Department of Computer Science
Virginia Tech
Blacksburg, VA, USA
shoosh@vt.edu

Asma Shakil*
School of Computer Science
University of Auckland
Auckland, New Zealand
asma.shakil@auckland.ac.nz

Steve Riddle*
School of Computing
Newcastle University
Newcastle upon Tyne, UK
steve.riddle@ncl.ac.uk

Ilknur Aydin
Computer Systems
Farmingdale State College
Farmingdale, NY, USA
aydini@farmingdale.edu

Nayla Nasir
Department of Software Engineering
Blekinge Institute of Technology
Karlskrona, Blekinge, Sweden
nayla.nasir@bth.se

Tejasvi Parupudi
School of Computing and Augmented Intelligence
Arizona State University
Tempe, AZ, USA
parupudi@asu.edu

Attiqa Rehman
Department of Computer Science
Norwegian University of Science and Technology
Trondheim, Norway
attiqa.rehman@ntnu.no

Michael James Scott
Faculty of Screen, Technology, and Performance
Falmouth University
Penryn, Cornwall, UK
michael.scott@falmouth.ac.uk

Jan Vahrenhold
Department of Computer Science
University of Münster
Münster, Germany
jan.vahrenhold@uni-muenster.de

Amali Weerasinghe
School of Computer and Mathematical Sciences
University of Adelaide
Adelaide, Australia
amali.weerasinghe@adelaide.edu.au

Xi Wu
School of Computer Science
University of Sydney
Sydney, NSW, Australia
xi.wu@sydney.edu.au

Abstract

Team-based capstone projects are vital in preparing computer science students for real-world work by developing teamwork, communication, and industry-relevant technical skills. Their assessment, however, is challenging, requiring alignment between academic criteria and external stakeholder expectations, fair evaluation of individual contributions, recognition of diverse skills, and clarity on external partners' involvement in the evaluation process. The high stakes of these projects further demand transparent and equitable assessment methods that are perceived as fair by all involved. Our working group (WG) addresses the challenges of capstone project assessment by examining the perspectives of instructors, students, and external stakeholders to support fair and effective evaluation. Building on insights from our previous WG and a comprehensive review of the literature, we used a mixed-methods approach combining online surveys (quantitative) and in-depth interviews (qualitative) with instructors, students, and external stakeholders. In total, we collected 66 survey responses and conducted 30 interviews across multiple countries and institutions, capturing a diverse range of global perspectives on capstone course assessments. Insights from instructors and students revealed several commonalities, for example, in the types of assessed components and the challenges of identifying and addressing non-contributing group members. Our findings also revealed clear variation between instructor and student perspectives on how contributions are measured and weighted. Instructors were reluctant to rely heavily on peer or self-evaluation due to concerns about reliability, preferring scaffolded assessments and early-warning systems to gather contribution data and moderate team dynamics. They viewed contribution-based grading as positive but resource-intensive. Students, in contrast, emphasized the need for more transparency, formative feedback, and accurate recognition of individual contributions. They also expressed concerns about the lack of recognition for *hidden labor* (e.g., project management, team coordination), assessor inconsistency, and a reluctance to critique peers. Instructors treated peer input as supplementary evidence, whereas students perceived it as high-stakes and socially risky. Stakeholder involvement in assessment was

*Co-leader



This work is licensed under a Creative Commons Attribution 4.0 International License.
ITiCSE-WGR 2025, Nijmegen, Netherlands
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2167-0/2025/06
https://doi.org/10.1145/3760545.3783974

generally limited to providing formative feedback and participating in final showcase events. We also identified generative AI as a rapidly evolving challenge, with both students and instructors seeking guidance on acceptable use and exploring opportunities to automate aspects of assessment. Our results offer actionable evidence-based guidance for designing transparent and equitable assessment practices in team-based computing capstones.

CCS Concepts

• **Social and professional topics → Computing education**.

Keywords

Capstone, Team-based Assessments, Individual Contribution

ACM Reference Format:
Sara Hooshangi, Asma Shakil, Steve Riddle, Ilknur Aydin, Nayla Nasir, Tejasvi Parupudi, Attiqa Rehman, Michael James Scott, Jan Vahrenhold, Amali Weerasinghe, and Xi Wu. 2025. Evaluating Assessment Practices in Team-Based Computing Capstone Projects. In *2025 Working Group Reports on Innovation and Technology in Computer Science Education (ITiCSE-WGR 2025), June 27-July 2, 2025, Nijmegen, Netherlands.* ACM, New York, NY, USA, 36 pages. https://doi.org/10.1145/3760545.3783974

1 Introduction

Team-based capstone projects mark the culmination of computer science training, allowing students to apply their knowledge and cultivate professional skills through real-world problem-solving [29, 50]. Fair and meaningful assessments in these courses ensure that students' individual contributions are recognized and learning outcomes are accurately measured [19]. However, instructors face challenges when assessing capstone projects, such as balancing subjective and objective criteria, managing team dynamics, and addressing diverse expectations [59]. Meanwhile, students often struggle with equitable contribution distribution, understanding assessment criteria, and receiving actionable feedback. Moreover, generative AI developments require instructors to adjust to the challenges and opportunities of responsible use of AI in areas such as assessment [46]. Our WG aims to explore the complexities of capstone-specific assessments and builds on findings from a previous WG that identified assessment as a significant challenge in capstone courses [23]. Unlike the broader teamwork evaluation studied by a SIGCSE 2024 WG [36], capstone-specific assessments require focused investigation as they present distinct challenges. These challenges include aligning academic criteria with professional standards and real-world deliverables, often necessitated by the involvement of external clients/ industry partners. Capstone projects require students to demonstrate diverse technical, managerial, and interpersonal skills, which must be evaluated consistently across varied contexts. The high stakes of capstone projects also require transparent and equitable assessments that all involved perceive as fair.

1.1 Motivation

While prior research has explored pedagogical models and student outcomes in computing capstones, there has been less attention paid to how assessment is experienced and implemented by the various groups involved, including students, instructors, and external clients. This gap provides the primary motivation for our current work. In an earlier study [23], we interviewed over 30 faculty members about their capstone teaching experiences, uncovering recurring concerns about assessment that we were not able to fully explore at the time. As instructors who have each faced the complexities of assessing group work firsthand, a secondary motivation for this study is to better understand these challenges across roles and institutions, and to surface practices that can support more equitable and effective evaluation in capstone settings. To this end, the current study examines assessment from multiple perspectives, explicitly including students and stakeholders — two groups that may not shape the design or deployment of assessment but whose perspectives are nevertheless crucial for building a more holistic understanding.

1.2 Research Questions

While many instructors use a variety of methods to assess student work in capstone courses, there is limited cross-institutional understanding of which approaches are most effective, how they are perceived by students and stakeholders, and what challenges instructors face in implementing them. This project aims to gather insights from published research and first-hand discussions with instructors, students, and stakeholders to explore assessment from multiple perspectives. Our goal is to gain a deeper understanding of current practices, identify persistent challenges, and pinpoint opportunities for improvement. Our goals translate into several research questions which are as follows:

- **RQ1:** What assessment methods are commonly used across diverse computing capstone courses, and how effective are they in evaluating both team and individual contributions?
- **RQ2:** What challenges do instructors face in assessing team-based capstone projects?
- **RQ3:** How do students perceive the fairness, transparency, and clarity of assessment criteria, feedback, and grading in capstone courses?
- **RQ4:** What roles do external stakeholders play in the assessment process of capstone projects?
- **RQ5:** What evidence-based strategies can address key assessment challenges while supporting both instructor and student needs?

2 Capstone Project Terminology

Members of our working group represent seven different countries and cultures spread across the globe, so the first step is to establish common terminology. In the context of this study:

- A **course** is an individual unit in a program of study. This is known as a module or class in some countries [6].
- **Capstone** refers to a team-based course completed in the final or penultimate year of study. Other countries use the term to describe an individual project carried out in the final year [23].
- An **instructor** is a faculty member (academic) who oversees the delivery and assessment of the capstone course. Other members of the faculty team might include course designers and assessors.

- A **stakeholder** is a role separate from the course, typically an industry partner, client or sponsor involved in the technology industry, serving as a provider of project specification, advisor, informal assessor or mentor for the capstone team.

These definitions and terminology are not universal. It is not unusual for individuals to have different roles, for example an instructor in one context may also act as a stakeholder at a different institution. We discuss the potential implications of the differing contexts associated with this work in Section 10.

3 Related Literature

In this section, we review the literature on assessment practices in capstone courses, examining commonly reported approaches, the challenges faced by instructors and students, and the involvement of external stakeholders. We begin by contextualizing our discussion by examining established assessment frameworks in general engineering capstone education. We then focus on computing-specific capstone assessments with particular attention to the evaluation of individual and team contributions, the roles of different assessors, and the perceived fairness and effectiveness of these assessment practices. Finally, we identify gaps in capstone assessment research, particularly around student and stakeholder perspectives.

3.1 Broader Context: Assessment Practices in General Engineering Capstone Education

Research in the broader engineering education community has long recognized capstone courses as a cornerstone for integrating learning and professional skill development. Foundational studies, such as McKenzie et al. [35] and Trevisan et al. [60], mapped assessment practices across institutions, revealing that while capstones are central to ABET[1] outcome assessment, approaches vary widely and tend to emphasise summative evaluation over formative learning. Subsequent work introduced more structured, transferable assessment tools. For example, Davis et al. [14] identified three core performance domains (personal capacity, team processes, and solution quality), while Chowdhury et al. [12] demonstrated the role of capstones in continuous program improvement through alignment with accreditation outcomes.

Several researchers have proposed a holistic assessment model. Nassersharif and Rousseau [39] developed a rubric-driven framework integrating student, instructor, and sponsor feedback to ensure consistency and transparency. Steiner et al. [53] further distinguished mentoring from grading and adopted multi-source evidence to improve fairness and reliability. Davis and Rogers [15] later extended this work through a stakeholder-centred structure that aligns capstone assessment with student, faculty, and employer priorities, operationalising evaluations around key knowledge, skills, and abilities (KSAs) such as communication, teamwork, project management, and ethics.

Collectively, this body of research positions engineering capstones as authentic, integrative assessments of professional competence, offering mature models for rubric design, multi-source evaluation, and accreditation-aligned feedback. However, these approaches remain grounded in physical product design and multidisciplinary engineering contexts. In computing and software engineering, assessment challenges differ, e.g., centred on intangible artefacts such as code, documentation, and iterative prototypes, and framed by agile, client-driven processes. Our review focuses on assessment practices within computing-related capstones, examining how they have evolved to address these distinct pedagogical and evaluative contexts.

[1]ABET Criteria for Accrediting Computing Programs. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2025-2026/

3.2 Assessment Practices in Computing Capstones

Prior studies of computing capstone courses describe a wide range of assessment approaches. We organize our findings around four analytic dimensions: methods employed, who assesses, artifacts and deliverables, and perceived effectiveness.

3.2.1 Methods Employed. This subsection focuses on the *assessment methods* used in computing capstones, that is, the procedures, processes, and mechanisms through which student work is evaluated, independent of the specific artifacts produced.

The literature reveals a consistent reliance on a blend of formative and summative approaches, often combining deliverables-based grading, peer and self-assessment, process evaluation, stakeholder input, and quantitative analytics. Formative methods include weekly progress reviews, reflection reports, status updates, mentor check-ins, client feedback, and sprint reviews, which provide ongoing guidance and opportunities for teams to adjust their processes [3, 5, 9, 10, 64]. Summative methods typically comprise graded demos, public presentations, final reports, and the final product, with assessments conducted at multiple stages of the project life cycle [25, 42, 51, 57]. Some courses introduce structured checkpoints or milestone reviews to keep teams on track [17, 25, 28], while others use manual grading augmented with automated metrics from tools such as Agilefant to evaluate progress [43].

Peer- and self-assessment are widely used to individualize marks, address free riding, and make team dynamics more visible. Approaches range from simple Likert-scale questionnaires to mathematically weighted schemes (e.g., a three-dimensional sigmoid combining group performance, individual effort, and peer assessment average) [4]. Some courses co-define assessment criteria with students to improve buy-in and alignment with learning objectives [65]. Most implementations require structured, often anonymised, peer reviews and reflective self-reports that inform adjustments to individual marks [9, 48, 51, 56]. These processes are scaffolded by clear rubrics and digital tools to standardize contribution measurement and promote fairness [21, 50, 61, 63]. In many cases, peer evaluations contribute directly to the summative grade [4, 28].

Stakeholder and client feedback is also a common component, with industry mentors or clients evaluating intermediate and final results, often providing formative input and end-of-course judgment on the project's success [3, 5, 40]. Increasingly, programs integrate digital platforms and automated tracking tools into the assessment process. For example, GitHub analytics can capture code commits, pull requests (PRs), and issue tracking to objectively evaluate contributions [21], while tools like Git-Truck provide visual analytics of repository activity to support assessment of workload distribution and progression [41]. In other cases, deliverables are

weighted across quizzes, reports, and final technical papers as part of a comprehensive grading scheme [58].

Table 1 summarizes the key assessment methods identified in this subsection, providing a concise overview of the major approaches reported in the literature.

Table 1: Summary of Assessment Methods Reported in Computing Capstones

Method Category	Examples in the Literature
Formative Assessment	Weekly check-ins, client/mentor meetings, sprint reviews, reflection reports, status updates [3, 5, 9, 10, 64]
Summative Assessment	Final demos, presentations, reports, product evaluations, milestone reviews [17, 25, 28, 42, 51, 57]
Peer / Self Assessment	Likert-scale peer reviews, weighted grading schemes, reflective self-reports [4, 9, 28, 48, 51, 56, 65]
Stakeholder Input	Client evaluations, industry mentor feedback, communication skill ratings [3, 5, 40]
Digital Analytics	GitHub metrics (commits, PRs), Agilefant statistics, Git-Truck repository analytics [21, 30, 41, 43]

3.2.2 Assessors. Assessment in capstone courses is typically carried out by several stakeholders whose roles either overlap or complement one another. Instructors and teaching assistants (TAs) remain the principal arbiters of technical quality and teamwork, commonly determining final grades while also providing formative feedback [3, 24, 41, 57, 69]. Additional studies report the same central role for teaching staff across iterations and deliverables [4, 22, 50, 55, 64]. In some instances, instructors are the sole assessors of student work [10, 25, 67]. In others, they share responsibility with clients, sponsors, or students through mixed evidence and rubrics [5, 17, 57, 69]. Large-scale or mentor-led formats may involve many academics supervising and assessing teams [41]. Faculty or academic mentors can also be assigned per team for continuous evaluation [42, 58, 62]. Teaching assistants and tutors frequently act as assessors, agile mentors, or Scrum coaches. They review process and product, provide ongoing feedback, and sometimes even play the role of customers to articulate requirements [30, 33, 65, 66]. Tutors and graduate assistants also contribute to formative assessment and help enforce consistency across teams [5, 11]. Students may also hold explicit assessor roles (e.g., team leaders evaluating contributions) or facilitate reflective assessment in collaboration with staff [34, 68].

External clients and stakeholders provide authentic, real-world perspectives. In many offerings, their role is mainly formative, such as commenting on demonstrations or final presentations and validating the product's value [4, 51, 69]. In other courses, clients contribute substantially to summative assessment, sometimes accounting for up to half of the final grade by judging usefulness and professionalism [3, 11, 52]. In industry-integrated capstone models, industrial mentors, or project supervisors actively evaluate deliverables and guide work across iterations [30, 34, 65]. In a few cases, client approval is decisive for success, with acceptance or rejection of the delivered system determining the outcome [40, 44].

Some programs employ coordinators, supervisors, or formal boards to balance perspectives and ensure consistent standards. A single coordinator or academic supervisor may oversee all teams and moderate grading across iterations [22, 61]. Formal boards or committees comprising faculty mentors, clients, and additional instructors jointly evaluate final outcomes [42].

Table 2 provides a summary of the primary assessor roles in computing capstones, consolidating the diverse perspectives highlighted in this subsection.

Table 2: Assessors in Computing Capstones

Assessor Category	Roles and Contributions
Instructors / TAs	Technical evaluation, process supervision, formative feedback, final grading [3, 4, 10, 22, 24, 25, 41, 50, 55, 57, 64, 67, 69]
External Stakeholders	Domain-fit evaluation, professionalism, product usefulness, communication skills [3, 4, 11, 51, 52, 69]
Students (Peer / Self)	Contribution ratings, team dynamics insight, reflective assessment [34, 68]
Mentors / Coaches / Supervisors	Regular progress evaluation, process guidance, intermediate deliverable assessment [30, 34, 65]
Committees / Boards	Moderation, multi-source final evaluation, ensuring consistency across teams [22, 42, 61]

3.2.3 Artifacts and Deliverables. This subsection focuses on the computing-specific *artifacts and deliverables* that are assessed in capstone courses, distinguishing what students produce (e.g., code, models, documentation) from the assessment methods described in section 3.2.1.

Core graded artifacts include software increments and final products, requirements and design documents, and technical implementation records [4, 5, 10, 40, 41, 48, 50, 54, 55]. These deliverables are often weighted alongside measures of process and teamwork, with some courses using metric-based schemes to evaluate productivity, teamwork quality, and product quality [67]. Oral presentations and demonstrations are common at milestones and final delivery [10, 11, 22, 25, 51], while written reports, such as proposals, interim updates, and final documentation, remain central [9, 17, 38, 57, 61]. Technical artifacts assessed include use case models, test models, architecture diagrams, and prototypes [30, 34, 56, 66] while process evidence artifacts include sprint backlogs, burn-down charts, retrospectives, meeting logs, and contribution breakdowns [28, 61, 64].

Reflective deliverables, such as learning journals, self-appraisals, and iteration retrospectives, support meta-cognitive learning [3, 31, 62]. Increasingly, code repositories are assessed for quality and contribution, with analytics tools like Git-Truck used to visualize team and individual inputs [21, 41]. In multidisciplinary or

industry-linked projects, additional outputs include posters, electronic team notebooks, and client-approved final designs [2, 38]. Structured models such as the Triangulation Assessment Model integrate rubrics, logbooks, and surveys to evaluate both product and process [20].

Agile and iterative project settings emphasize "living documents" such as requirements specifications, test documentation, architecture diagrams, Epics, sprint outputs, and CONOPS (Concept of Operations) documents [30]. Other technical artifacts include 3D models, UI components, functional prototypes, implementation logs, and function point estimation reports [8, 34, 56]. Some curricula incorporate assessment-oriented artifacts such as progress reports, contribution breakdowns, and formal evaluation documents like project notebooks, risk management plans, and traceability matrices [42, 65]. Evaluation schemes may also include conference-style technical papers, engineering journals, and database logs of process adherence [65].

Table 3 summarises the categories of artifacts and deliverables that are assessed in computing capstone projects.

Table 3: Summary of Artifacts and Deliverables Assessed in Computing Capstones

Artifact Category	Examples in the Literature
Software Products & Increments	Functional prototypes, final software builds, incremental releases, implementation records [4, 5, 10, 40, 41, 48, 50, 54, 55]
Technical Documentation	Requirements specifications, design documents, architecture diagrams, UML/use-case models, test models, CONOPS documents [30, 34, 56, 66]
Written Reports	Proposals, interim reports, final documentation, engineering journals, conference-style papers, evaluation documents [9, 17, 38, 57, 61, 65]
Oral Presentations & Demonstrations	Milestone demos, technical presentations, final project showcases, poster presentations [10, 11, 22, 25, 51]
Process Evidence	Sprint backlogs, burndown charts, retrospectives, meeting logs, risk management plans, traceability matrices [28, 42, 61, 64, 65]
Code Repositories and Analytics	Git commits, merge requests, contribution visualisations (e.g., Git-Truck), repository activity metrics [21, 41]
Reflections	Learning journals, self-appraisals, iteration retrospectives [3, 31, 62]
Other Deliverables	Client-approved designs, 3D models, UI components, function point estimation reports, project notebooks [8, 34, 42, 56, 65]

3.2.4 Perceived Effectiveness. Courses that incorporate multiple sources of evidence, combining inputs from instructors, self, peers, and clients, tend to report higher perceptions of fairness and transparency [3, 5, 38, 68].

In practice, instructors frequently *triangulate* these data streams by integrating peer evaluations, customer feedback, and their own judgment when determining grades. They may also consult team leaders to deepen insight into individual contributions during self- and peer-assessment cycles [58]. In addition, timely and actionable feedback from assessors improves the perceived effectiveness of assessment [62].

Some studies describe concrete measures to enhance fairness and coordination. For example, *Reactive Weekly Monitoring* (RWM) has been shown to improve team coordination, effectiveness, and sense of belonging (though not necessarily productivity) [34]. To further strengthen objectivity, some instructors propose automated collection and analysis of repository and development-activity data to complement individual performance assessment [21].

Despite these benefits, some authors caution that peer and/or self-ratings can suffer from friendship bias or strategic inflation [5, 48]. From the instructors' perspectives, large class sizes and different assessor viewpoints complicate calibration and increase workload [24, 62]. Client-weighted schemes can also become unreliable when stakeholder availability is uneven or feedback is misaligned with learning objectives [57].

Notably, evidence on the effectiveness of assessment practices in computing capstones is limited and often indirect. While studies frequently report that certain assessment components, such as iterative feedback, multi-source evaluation, or structured rubrics, are positively received by instructors or students, few provide systematic measures of effectiveness or formal evaluations of how well these practices differentiate individual from team performance. This scarcity of explicit effectiveness measures reflects a broader gap in the literature and directly motivates the second component of RQ1, which asks how well current methods evaluate both team and individual performance.

3.3 Instructor Challenges

Our review of the literature reveals that capstone course instructors face a wide range of persistent and emerging challenges, including managing limited time and resources, navigating diverse teams and stakeholders, assessing individual contributions, balancing process- versus product-oriented criteria, coordinating multi-source feedback, and contending with non-standardized assessment models. These challenges are used to structure our findings, below.

None of the studies included in our review considered generative AI tools for assessment tasks. This leaves open questions about how to detect AI generated code and embed AI literacy in capstone assessment rubrics. However, as will be seen from the later sections, interviews with instructors, students and stakeholders demonstrate evolving practice around AI.

3.3.1 Workload, Time Constraints, and Resources. A commonly reported issue in the literature is the lack of time to adequately monitor and assess both individual and group-level progress [57, 66]. Providing timely, detailed, and high-quality individual feedback is described as "extremely time-consuming" [22, 61]. Adding to this, instructors report that some students do not act on feedback,

leading to repeated under-performance and duplicated effort from teaching teams [61].

Resource constraints also present barriers to the implementation and evaluation of innovative assessment techniques. For example, the use of Git-Truck to assess student contributions required 19 teaching staff for 19 student projects, effectively a one-to-one ratio between staff and project, making such models infeasible for most institutions [41].

The limited number of academic mentors available for a large student cohort is also a challenge [32]. This also complicates the task of effectively matching students, projects, and mentors. To provide meaningful technical guidance, instructors must invest substantial time in understanding each project while also managing course design, scheduling, and assessment planning, each of which is resource-intensive [55].

Large enrollments in capstone courses significantly increase the grading workload. For example, Linköping's large-project course required a ten-activity marking model and multiple assessors to manage ninety students [63]. Similarly, frequent sprint reviews in client-on-site settings can stretch instructor capacity [3, 49].

3.3.2 Diversity of Projects, Teams, Instructors, and Stakeholders. In capstone project courses, assessment is more complex than in other courses due to the inherent multi-dimensional nature of capstone projects, the entanglement of generic skills (like problem formulation and critical thinking), and the need for continuous evaluation[2]. Since capstone courses have a variety of projects [45], and a diverse set of students [51], it is challenging to establish a uniform assessment process. Some courses [69] use an assessment model where each team works on the same project to create a consistent and fair base for grading, as each team deals with the same project complexity, management, and technology issues. In some institutes, capstone courses are offered to both Bachelor's and Master's students in their final year of study, with 30 vs. 40 hours of load per week, leaving varying availability for face-to-face study on a daily basis [51]. Conflicts happen within a team due to the coexistence of certain personality types [48].

Further challenges for instructors come in relation to the industry. For example, managing projects with inconsistent levels of external mentorship from stakeholders [24] requires reconciling different evaluation perspectives while assessing a broad range of skills and outcomes. Variations in management and assessment practices make it harder to align student capabilities with industry demands [2].

Further variations arise in differences in opinion of diverse assessors in terms of quality and ethical consideration [24, 38, 61] and differing experiences of instructors influencing the design of problem-based learning (PBL) assessments [68].

3.3.3 Tracking Individual Contribution. Objective measurement of individual student engagement within a team [48] and detection of free-riders are challenging tasks for instructors [4, 11, 18, 48] especially when team members are reluctant to report problems with their teammates until is too late to fix the issues, and when assessments start late in the semester [25]. The teaching teams largely rely on students' personal reflections and their own observations, which could be limited or biased [50]. Identifying and addressing so-called "social loafing" or free-riding is also difficult because it could go undetected for a long time, and by the time it is reported, the course may be almost over [50].

When projects are significantly different, consistent and accurate assessment of individual student contributions to team-produced work presents a bigger challenge to instructors [4, 22, 50]. Providing early feedback helps students succeed [64]. Instructors utilise tools such as Trello, Git repositories, and shared time-logging spreadsheets to gain insights into individual contributions [30, 56]. Instructors also often rely on documentation artifacts (e.g., Project Notebooks) and weekly meeting notes to infer each student's contribution [58]. Tools such as the Team Contribution System (TCS) provide numeric peer scores, but bias and strategic inflation remain [5]. Automated log metrics (e.g., commit counts) help, but require additional analytics expertise [28].

3.3.4 Impact of Team Setup on Assessment. Team configuration also shapes the assessment challenges that instructors encounter. Hybrid, virtual, and distributed teams reduce instructors' opportunities for direct observation, making it harder to monitor engagement, identify problems early, and verify consistent participation. Studies report that remote or dispersed teams experience weaker visibility of day-to-day work, greater reliance on asynchronous communication, and higher coordination overhead, often leading to delayed issue reporting and difficulty establishing shared norms [1, 47, 58].

Because of these constraints, instructors must rely more heavily on digital traces, such as commit histories, issue trackers, and meeting logs, to infer contributions. However, these artifacts capture only part of the work and require careful interpretation [28]. These limitations intensify individual contribution-tracking issues and highlight the need for assessment approaches that remain reliable across co-located, hybrid, and fully distributed team setups.

3.3.5 Balancing Process–vs–Product Criteria. Assessing the capstone project process versus the product in a balanced manner is a complex task. While some grading schemes allocate weights to product quality, process deployment, and teamwork (e.g., 40%-30%-30%) [58], other approaches prioritize process adherence more explicitly, such as assessments based on CMMI maturity levels to evaluate the depth and structure of team practices [65]. Agile-oriented offerings aim to reward professional practice, not just the final artifact. For example, Eloe and Hoot caps MVP marks at "Distinction" unless the process evidence shows a higher complexity and reflection [17]. Other approaches seek to reward learning and not just outcomes [25] by taking into account group skills, project complexity, and standard of outcome, so that a team's ability to solve challenges and learn from them is rewarded. Overall, achieving consensus on such weighting schemes across teaching teams is non-trivial.

3.3.6 Coordinating Multi-Source Feedback. The need to ensure objectivity and integrity of the final grade by correlating grades from a variety of sources is also reported as a challenge by instructors [22]. Since differences in scores of different assessors were reported as a concern of students [68], instructors of the capstone project course need to be extra careful when coordinating multi-source feedback. Some studies reported uneven client availability when students needed feedback [48]. Instructors thus act as liaisons, interpreting stakeholder comments and converting them into actionable student feedback. It is difficult for instructors to measure

and determine if the intended outcomes for all parties are achieved. Monitoring the interaction and communication between students and external stakeholders is not easy, which makes it harder to ensure the achievement of learning objectives and project goals [40, 52]. Instructors also face disappointment when students don't follow assessment instructions [55], when integrating the external stakeholders isn't perceived as "worth the effort", and when course evaluation is poor because the external stakeholder didn't meet the student needs [52].

3.3.7 Non-Standardized Assessment Models/Frameworks. Due to the heterogeneity of capstone projects in terms of their types, subjects, end-users, and teams' capabilities in developing them, it becomes very difficult to get a standard measure of their quality and have a standardized assessment criterion [20]. The lack of standardized assessment frameworks for the variety of capstone projects leads to inconsistent implementations due to the varying approaches of educators, institutional factors, and challenges in maintaining consistency between instructors [20, 62].

3.4 Student Perspectives

Capstone students perceive various challenges, including unclear assessment criteria, a lack of fairness, complexities with the peer and stakeholder assessments, and issues with timeliness and the usefulness of the feedback. These challenges are expanded on in the following sections.

3.4.1 Transparency and Clarity of Criteria. Students prefer up-front rubrics and briefing sessions to help them understand the assessment process in capstone projects. Studies show that ambiguous weighting between process and product undermines student confidence [57]. Students are affected by the inflexibility in distributing the work and marks between project phases and in deciding group sizes, which may impact their learning experience [2]. Moreover, discrepancy in scores of different assessors is reported as a student concern [68]. Instructors rarely request student feedback on specific aspects of capstone assessment (such as peer assessment as a method of individual assessment, effects of friendship on peer assessment, evaluation of the peer assessment form, peer assessment grade reflecting minimum effort, and evaluation of the adjustment formula) [4].

3.4.2 Peer Assessment Impact. When used formatively, peer reviews promote accountability and reflection [18]. Fairness in peer evaluation is emphasized, with students encouraged to assess teammates honestly and objectively [30]. Measures are taken to prevent bias, such as disregarding self-evaluations that appear overly inflated, and peer assessments are used to adjust final grades, which adds credibility to the evaluation process [58]. However, fear of confrontation of friends results in inflated ratings, diluting its diagnostic value [5]. Students may even feel uncomfortable in assessing their friends [4]. The lack of student expertise in assessment leads to high scores in self-assessment as compared to instructor or other assessor scores [20]. Furthermore, an extrovert "free rider" may generate favorable assessment in the peer evaluation based on their personalities, while introvert students might be marked down [4].

Complaints about unequal effort persist, particularly when peer scores are introduced late in the course [5]. Hybrid grading (group + individual) is preferred by a majority of students, such that 68% of students in one study rated it fairer than pure team marks [63]. However, some studies report that students demanding fairness in grading sometimes overestimate their ability to achieve a high grade [62].

3.4.3 Working with Stakeholders. Involvement of the stakeholders in the capstone projects may cause mixed responses in students. For example, students can be frustrated when the final product is not part of their assessment, and hesitant to work with the external stakeholders on an open problem - but better engaged in giving demonstrable results to an external product owner on a weekly basis [9]. Students encounter other common problems while working with the stakeholders, including absentee clients, a lack of client-supplied content or data, and excessive scope creep of the project, which may lead to project failure and impact the quality of the final software product students are assessed on [22, 52]. Students also experience frustration due to different terminologies used between the external stakeholders and instructors, and face cognitive overload when working with open problems, new tools, processes, and external stakeholders simultaneously [52]. Students also report difficulty in balancing the course requirements with external stakeholder demands, being sometimes confused by conflicting advice from external stakeholders or excessive focus on the final product over learning practices and learning outcomes [52].

3.4.4 Timeliness and Usefulness of Feedback. Students value weekly tutor meetings and sprint retrospectives as the most actionable feedback channels [28]. Summative comments delivered only at the end of the term were deemed "too late to matter". Sometimes, a lack of unified structure, documentation methods, and effective knowledge management leads to confusion, knowledge gaps, and integration problems. Under-communication during times of uncertainty is especially challenging to handle by the students and can easily cause students to lose the "big picture" of the whole project [50].

3.4.5 Other Challenges Impacting Assessment. For many students, the capstone is their first experience with a large-scale, real-world project, and they report difficulties in knowing how to begin, especially within the tight timelines typical of these courses [55]. Meeting deadlines can be especially problematic when sub-teams fall behind, as this jeopardizes overall project integration and impacts assessed outcomes [50]. Students also describe anxiety at the outset of the course when forming teams or being allocated projects, with mismatches in project interest sometimes contributing to reduced motivation and weaker performance [2]. Studies also show that students often overestimate the usability of their technical solutions, assuming users will intuitively understand how to interact with a system [8]. Feedback from authentic users, such as teachers or peers, helps them recognize usability shortcomings and motivates design improvements, for example, by adding hint buttons or visual guidance [8].

Teamwork itself is a recurring source of difficulty. Students acknowledge that project planning and collaboration are more demanding than anticipated, and they often underestimate the complexity of team dynamics before the course begins [5]. Conflicts, uneven contributions, and coordination issues can lead to poor deliverables and lower grades [22]. Even when digital platforms

for communication and management are available, maintaining coordination under strict time constraints remains challenging [2]. Broader systemic issues, such as unclear work distribution, limited knowledge sharing, and concerns about assessment schemes, further complicate students' experiences and their perceptions of fairness in grading.

3.5 Stakeholder Involvement

Our review of the literature shows that the involvement of stakeholders in capstone courses is limited, but varies significantly. The sections below discuss the nature of stakeholder involvement, their input to the assessment, feedback, and expectations.

3.5.1 Extent and Nature of Involvement. External stakeholders, such as clients and end users, can be part of both formative and summative assessment processes. Their involvement can range from limited engagement, such as serving on a final judging panel, to sustained participation throughout the project, including co-defining requirements, attending sprint reviews, and providing iterative feedback [3, 38, 51, 57]. Across the literature, stakeholders contribute anywhere from a small portion to a substantial share of the overall assessment, typically falling within a broad range from light weighting to nearly half of the total evaluation [24, 51].

Stakeholders contribute to various phases of the project life cycle, providing formative feedback that helps students improve their systems iteratively. For instance, feedback from system users guides enhancements in usability and functionality [8]. Project ideas can be solicited from the industry partners, but not always used as they were originally proposed and instead adjusted for the course requirements [69]. While the instructor is the primary assessor of the quality of the work, industry partners can provide input on the acceptability of the project implementation and deployment [69]. In another example, external sponsors meet with the students weekly to provide feedback on the progress along with the instructor (the main assessor) and other teaching staff and postdoctoral students [25]. In some cases, assessment models also incorporate stakeholder input to ensure a holistic evaluation. For example, grades on team deliverables are first determined at the group level and then adjusted based on individual contributions as judged by the instructor, project customer(s), and peers [58]. In other cases, client involvement extends beyond the course, offering post-project employment opportunities to students [42].

Stakeholders evaluate student performance in terms of professional competency, the gap between academic and industry standards, specific professional skills (including drawing, modeling, coordination, cost estimation, and teamwork), capstone achievements, and overall post-graduation readiness [32]. Conflicting schedules, shifting priorities, and discrepancies in quality standards can lead to misaligned expectations [38, 63]. To mitigate these issues, courses may assign dedicated mentors, employ common rubrics to standardize evaluation [38], or require teams to follow clearly defined processes, such as structured Agile practices, that help align stakeholder expectations with course outcomes [28].

3.5.2 Input to Assessment, Feedback, and Expectations. External stakeholders contribute to the capstone assessment by evaluating usability, domain fit, and professional conduct, and their qualitative feedback often informs instructors' grading [5]. Some programs involve industry guests in rating communication skills during milestone demos or in applying stakeholder-provided rubrics to assess final designs, posters, and presentation quality [24, 51, 57]. Stakeholder evaluations may therefore form a direct component of individual student assessment.

Beyond grading, stakeholders provide insight into professional competencies, industry expectations, and perceived gaps between academic preparation and workplace standards [32]. However, differences in scheduling constraints, shifting priorities, and varying quality benchmarks can lead to misalignment between academic and industry evaluations. For example, industry mentors often show greater variability in their evaluations and may assess performance differently from academic staff, reflecting divergent expectations rather than consistent grade deflation [38].

3.6 Synthesis and Research Gaps

Our review of the literature shows that while capstone assessment research has explored what is evaluated (e.g., deliverables, professional skills), how it is evaluated (e.g., rubrics, multi-source grading), and who is involved (e.g., instructors, peers, clients, students), several important gaps remain. Assessment practices, such as checkpoint-based grading, triangulated reviews of code, documentation, and presentations, and multi-source input combining instructor, peer, self, and client perspectives, are reported, yet their adoption is inconsistent due to the substantial time, coordination, and resources required. Scalable, objective methods for assessing individual contributions within large teams remain scarce, and no studies were found that address the emerging challenge of recognizing and appropriately crediting generative AI contributions. Moreover, student perspectives are under-represented: few studies investigate perceptions of rubric clarity, feedback timeliness, or the fairness of peer assessment. The role of external stakeholders in assessment is similarly under-examined.

These gaps raise several key questions: How can instructors balance individual and team contribution and incorporate feedback from stakeholders while maintaining fair, fine-grained assessment? How do students interpret fairness, feedback, and rubric clarity? Addressing these questions requires research that not only examines the practical constraints faced by instructors but also integrates student and stakeholder perspectives to ensure capstone assessment practices are equitable, effective, and relevant. Linking back to our research questions, existing work offers partial insight into RQ1, particularly the range of assessment methods used, and addresses some aspects of RQ2 concerning instructor challenges. However, the literature provides limited evidence on how effectively these methods distinguish between individual and team performance (RQ1, second component), and offers little empirical understanding of students' perceptions of fairness, transparency, and clarity (RQ3). Research examining the roles and contributions of external stakeholders (RQ4) is similarly sparse. Finally, the absence of scalable, evidence-based strategies for addressing emerging assessment challenges, including issues introduced by generative AI, highlights a gap directly relevant to RQ5. These gaps collectively motivate the empirical components of our working-group investigation.

4 Theoretical Framework

This study is grounded in Social Interdependence Theory (SIT), which posits that the structure of group goals shapes member interactions and, ultimately, group outcomes [16, 26]. SIT distinguishes between three primary forms of goal interdependence: cooperative, competitive, and individualistic. In cooperative structures, which are ideal for most educational team settings, individuals work together to achieve shared goals, recognizing that their success is mutually dependent. Capstone courses typically rely on cooperative interdependence, as students must work collectively to complete complex, often externally sponsored projects. However, the nature of assessment in these contexts can create tensions. Peer evaluations, uneven contributions, competitive grading pressures, and unclear individual accountability can disrupt cooperative dynamics and shift perceptions toward competitive or individualistic structures. By applying SIT, we seek to understand how assessment practices either reinforce or undermine cooperative interdependence within capstone teams. This lens allows us to explore how students perceive fairness and contribution, how instructors manage group dynamics through assessment, and how stakeholder involvement influences shared goals and responsibility. Ultimately, this framework helps illuminate the alignment (or misalignment) between capstone pedagogy and the assessment structures that support or hinder effective collaboration.

5 Methods

Our work is guided by a mixed methods approach as outlined by John and David Creswell [13] to understand assessment from the point of view of various stakeholders involved with capstone courses. We collected information from students, instructors, and external stakeholders who acted as either consultants or project owners. We used a concurrent triangulation approach [13], in which quantitative and qualitative data are concurrently collected and then compared to determine convergence, differences, or some combination. In this approach, also referred to as cross-validation, *"separate quantitative and qualitative methods are used to offset the weaknesses inherent within one method with the strengths of the other (or conversely, the strength of one adds to the strength of the other)".*

5.1 Survey Design Process

We designed the survey and interview questions to align with our study's research questions and ensure the collection of relevant data, such as assessment type, fairness, followed by personal reflection on grading, individual versus group assessment, and the role of external stakeholders. The design of our survey and interview questions was guided by Social Interdependence Theory, previously introduced, to ensure that our assessment of computer science capstone courses reflects the central role of collaboration, interdependence, and shared goals in student learning. Two surveys were developed to collect quantitative responses from instructors and students. Three interview protocols were used for our semi-structured interviews to collect qualitative responses. Since this study involved 10 researchers from different countries and institutions, we developed guides, questions (Tables 15, 16, and 17), and a detailed interview protocol to ensure that the same procedures were followed at each site.

5.2 Data Collection Process

We distributed the quantitative survey (Tables 13 and 14) via a Google form to collect information from the larger community of faculty in the computing field who teach or supervise capstone projects, and to students at various institutions who have taken a capstone class. The surveys were distributed to the SIGCSE listserv and colleagues of members of the WG to share with their students. The results were collected and anonymized by the lead institution, which had the Institutional Review Board (IRB) approval for collecting the survey results before sharing with the rest of the research team. We collected 22 responses from instructors and 44 from students as shown in Table 4. While originally we planned to also survey stakeholders, unfortunately our call and outreach did not result in any response.

Simultaneously, semi-structured interviews were conducted by each member of the WG. Each researcher obtained IRB or ethics approval from their own institution and conducted the interviews in isolation. Each researcher conducted between 2 and 4 interviews. The participants were recruited from the personal network of each researcher. All interviews were conducted online using either Google Meet, Zoom, or Microsoft Teams. The transcripts were generated using the freely available automated audio transcription on Meet, Zoom, or Teams. The generated text was then cleaned and reviewed by the researcher for accuracy. While recruiting participants, we used a shared document to keep track of anonymized information about the participants.

Table 4 shows participant numbers in our study, divided into students, instructors, and stakeholders. There were 66 participants who responded to our quantitative survey and 30 who took part in the semi-structured interviews, resulting in a total of 96 participants across both sets.

Table 4: Survey/interview participants distribution

Role	Survey	Interview	Total
Student	44	12	56
Instructor	22	13	35
Stakeholder	0	5	5
Total	66	30	96

5.3 Coding Process

Most members of the research team conducted two to four interviews focusing on a single participant group (instructors, students, or stakeholders). With the exception of the lead authors, all researchers limited their interviews to just one group to maintain consistency. The research team then formed subgroups based on the type of participants each researcher interviewed—for example, those who interviewed students coded student transcripts, while those who interviewed instructors or stakeholders coded within their respective groups.

We analyzed the transcripts using thematic analysis, following Braun and Clarke's framework [7]. Given the clear separation of the three sets of interviews, each subgroup carried out its own thematic analysis. Each researchers in the subgroup independently coded their transcript, for an initial round of coding, then shared and

compared their coding with others in the same subgroup to refine interpretations and reach consensus. A code book was emerged from these discussions. To analyze the student interviews, the researchers deliberately distributed them in a way that interviews done at a particular institution were not analyzed by the researcher from that institution only. This way the team attempted to counterbalance threats arising from inferences or assumptions based on background knowledge about a particular institution. Finally, during the days prior to the ITiCSE conference, the researchers met in person to re-visit their codings, finalize the description of the emerging themes, and select the quotations used to illustrate the themes. This iterative process allowed each subgroup to surface patterns specific to their participant category, which resulted in the themes discussed in Sections 6.2, 7.2 and 8.

6 Instructors' Perspectives

This section presents insights from instructors gathered through both the online survey and in-depth interviews. Because some questions differed across the two instruments, we report the findings separately: quantitative survey data provide the breadth of perspectives across a larger group of instructors, while qualitative interview data offer depth of insight into individual experiences. The qualitative findings are organized around themes that emerged from the thematic analysis of the interview data. Collectively, these perspectives shed light on what forms of assessments instructors use and the strategies and challenges they face when assessing capstone projects.

6.1 Instructors' Quantitative Responses

6.1.1 Instructor and Course Background. Twenty-two instructors completed the online survey. Of those, 20 taught a capstone class at the undergraduate level and two at the graduate level. Eleven participants taught a one-semester capstone course, and eleven taught a two-semester or longer capstone. Most of the instructors were based in the United States (15), followed by the United Kingdom (3) and Norway (4). Most instructors were in a computer science department (18), and the rest were in gaming or informatics units.

For teaching experience, only one instructor had taught capstone courses for less than two years. Seven reported 2–5 years of experience, four reported 6–10 years, and ten instructors had been teaching capstones for more than a decade. The majority of the respondents were either a course coordinator or an instructor/teacher. Most instructors (13 out of 22) reported handling grading independently, while others noted involvement from other instructors, external stakeholders, and teaching assistants. When asked to rate the adequacy of grading support for assessing capstone projects, most instructors provided positive responses: 10 rated support as 'adequate' and 6 as 'extremely adequate', while the rest gave a more moderate rating of 'neutral'.

The class sizes ranged from less than 20 to more than 200. We asked instructors about the typical student team sizes in their capstone course, and the majority reported that student team sizes were between 3–5 (13 out of 22), followed by sizes greater than 6 (6 out of 22). These are the typical team sizes reported in the literature as well [59], as well as our own previous report [23].

6.1.2 Assessment Practices. We asked instructors about the form of assessments they used in their courses. Our results (Table 5) show that very few instructors rely exclusively on formative (ongoing) assessment, with no instructor in our sample size using it as the sole approach. Instead, most instructors adopt a combination of formative and summative assessment (19 out of 22), suggesting that instructors value both ongoing feedback and end-of-course evaluation. A similar pattern appears in assessment style: while some instructors employ exclusively subjective (3) or objective (4) measures, the majority (15) integrate both. Taken together, these findings indicate that capstone instructors blend multiple approaches, reinforcing the importance of capturing both process and outcome, as well as balancing qualitative judgments with more structured evaluation criteria.

Table 5: Assessment approaches reported by instructors

Approach	Frequency (N=22) (%)
Formative (ongoing) only	0 (0.0%)
Summative (end-of-course) only	3 (13.6%)
Both formative and summative	19 (86.4%)
Subjective only	3 (13.6%)
Objective only	4 (18.2%)
Both subjective and objective	15 (68.2%)

When asked about grading approaches, most instructors (10) reported assigning individual grades based on a combination of instructor, stakeholders' input and team member evaluations. A smaller group (4) indicated that grades were determined solely by the instructor's assessment of individual contributions, while another 4 reported giving all team members the same grade. Only a few instructors (3 in total) described hybrid or conditional approaches, such as splitting grades between group and individual components or assigning equal grades unless clear evidence suggested otherwise.

6.1.3 What is assessed? Table 6 shows the most common deliverables contributing to the final grade included project demonstrations (e.g., live demos or walkthroughs), written reports (e.g., final documentation or technical reports), and oral presentations. In some cases, additional components such as reflection essays or journals, peer feedback, and team logs were also used.

Table 6: Instructor-reported deliverables in capstone courses

Deliverable	Frequency (N=22) (%)
Project demonstrations (e.g., live demo, walkthrough)	19 (86.4%)
Completed project build/deliverable	18 (81.8%)
Written reports (e.g., documentation,report)	18 (81.8%)
Oral presentations (e.g., progress presentation, final pitch, final presentation)	16 (72.7%)
Reflection essays or journals	13 (59.1%)
Progress meeting minutes or team logs	12 (54.5%)
Others	6 (27.3%)

6.1.4 Contribution of Different Assessments to the Final Grade. Figure 1 presents instructors' response to the question of how different assessment types contribute to students' final grades in capstone courses, using a five-point Likert scale (Not at All to Very Significant). Instructor evaluation was rated as the most influential factor, with 86% of respondents viewing it as significant or very significant. In contrast, self-assessment was generally rated low: while half of the respondents rated it as only moderately significant, more than 40% reported it contributed little or not at all. Peer evaluation showed a more mixed picture, with 41% rating it as not at all significant, but nearly one-quarter viewing it as significant. External stakeholder feedback was the least emphasized, with nearly two-thirds of respondents reporting that it did not contribute meaningfully to final grades. Overall, the findings highlight the dominant role of instructor evaluation, with comparatively limited weight placed on self, peer, or external stakeholder assessment.

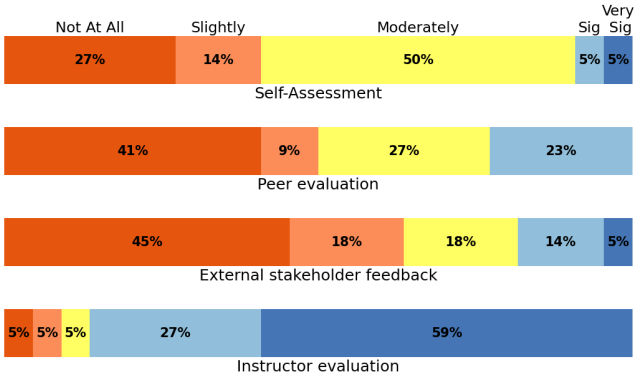


Figure 1: Extent by which each assessment type contributed to the students' final grades. Percentages may be 100% ± 1 due to rounding.

6.1.5 Evaluating Team and Individual Contribution. Figure 2 shows results for evaluating team contributions. Instructor evaluation again emerged as the most effective, with 86% of respondents rating it very or highly effective, while peer evaluation was viewed as moderately effective by nearly half, but less favorably by others. Self-assessment was perceived as moderately effective by 36%, but only 5% rating it as highly effective. External stakeholder evaluations were again the least used, with nearly half of respondents not employing them. These results suggest that instructor judgment and evaluation play the most central role in evaluating team-level contributions.

Figure 3 shows responses to the question of how effective different methods are in evaluating individual contributions in capstone courses. Instructor evaluation was rated the most effective, with nearly three-quarters of respondents (72%) considering it very or highly effective. Self-assessment received mixed reviews: while more than half (55%) considered it moderately effective, almost one-fifth did not use it. Peer evaluation showed the most divided opinions, with results spread across the scale; Over one-quarter of respondents (27%) regarded it as very effective, and a further 14% rated it highly effective, suggesting that some see peer insights

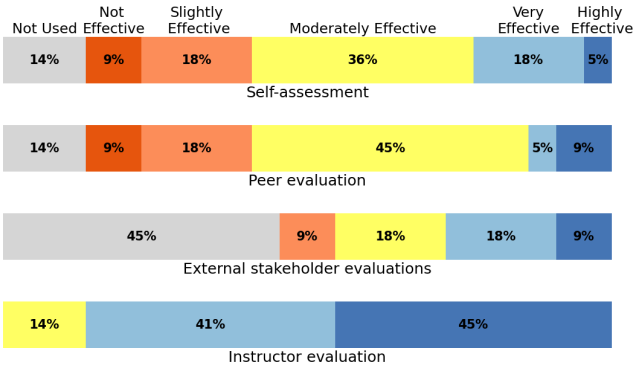


Figure 2: Effectiveness of different methods in evaluating team contributions in capstone courses. Percentages may be 100% ± 1 due to rounding.

as a valuable way of capturing individual contributions. At the same time, however, 23% judged it only moderately effective, 10% as slightly or not effective, and 27% reported not using it at all. External stakeholder evaluations were the least used, with nearly 60% of respondents not employing them for assessing individual contributions. Overall, the data indicate a strong reliance on instructor judgment, with other methods perceived as less consistently effective.

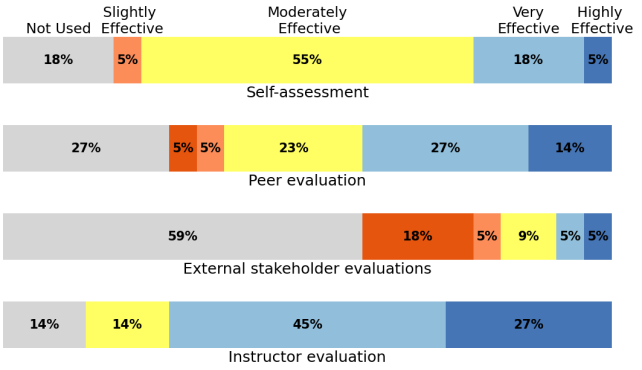


Figure 3: Effectiveness of different methods in evaluating individual contributions in capstone courses. Percentages may be 100% ± 1 due to rounding.

Taken together, the two sets of findings (Figures 2 and 3) highlight the similarities and differences in how individual and team contributions are assessed. Instructor evaluation stands out as the dominant method in both cases, underscoring its central role in capstone assessment. In contrast, self- and peer evaluations are viewed with more skepticism, especially for individual assessment, where concerns about reliability and subjectivity are more salient. Finally, external stakeholder evaluations remain the least utilized in both contexts. Overall, these findings suggest that capstone assessment practices lean heavily on instructor judgment, with limited reliance on alternative approaches.

6.1.6 Limitations and Challenges. Responses to the open-ended question about limitations of peer evaluations highlighted several

recurring concerns. Instructors noted that the quality of peer reviews was inconsistent, with students often awarding uniformly high marks regardless of actual performance. Some respondents observed that students could 'game the system' by informally agreeing to rate one another positively. Others emphasized that students were reluctant to 'throw their peers under the bus,' which contributed to inflated scores and limited the ability of peer evaluations to capture meaningful differences in individual contributions. Overall, these responses suggest that peer evaluations are often perceived as subjective and unreliable.

In response to the open-ended question, '*What limitations have you experienced in using external stakeholder evaluation for assessing individual and/or team contributions?*', several themes emerged. Respondents noted that feedback from stakeholders was often highly variable, sometimes vague, delayed, or incomplete, and frequently focused on the team as a whole rather than individual contributions. Concerns were also raised about stakeholders' limited ability to judge technical difficulty or student learning outcomes, with some tending to compare student teams to professional teams. In many cases, stakeholders were only involved in an advisory capacity, engaged late in the process, or not used at all, which further reduced their effectiveness in evaluation. Finally, respondents emphasized that because capstone courses are primarily learning experiences, stakeholder satisfaction does not always align with the educational goals of teamwork and skill development. Collectively, these observations underscore both the value of stakeholder perspectives and the challenges of relying on them as a central assessment tool.

When asked about the key challenges in assessing team-based projects, instructors identified several recurring difficulties as shown in Figure 4. The most common challenge was evaluating individual contributions, reported by 16 respondents. A closely related issue, noted by 15 participants, was the difficulty of allocating sufficient time for thorough evaluation. Bias in peer evaluations was also a concern, mentioned by nine respondents, while eight highlighted conflict within teams as a challenge that complicates fair assessment. Collectively, these findings underscore that distinguishing individual effort within team contexts and managing the practical and interpersonal complexities of group work remain the most pressing challenges in capstone assessment.

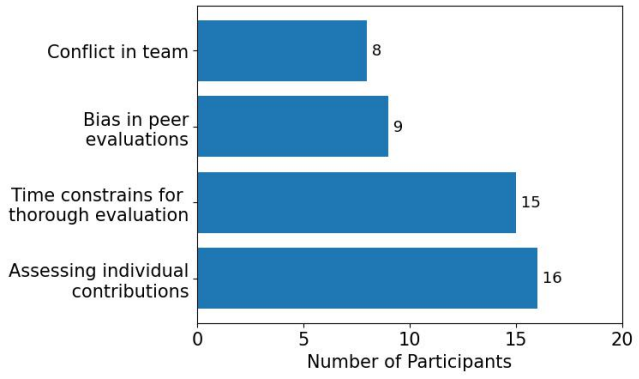


Figure 4: Key challenges encountered when assessing capstone projects

6.1.7 Use of Generative AI in Capstone Projects. Instructors reported varying levels of exposure to generative AI tools during the capstone process. Most had encountered their use at least occasionally (11 occasionally, 5 extensively), while only a few were unsure (3) or had not encountered them at all (3). This suggests that generative AI is already present in many projects, though often in limited ways.

When asked about the impact of generative AI on the overall quality of capstone projects, responses were mixed and largely neutral. Figure 5 presents the responses. Eleven respondents selected the neutral option, while smaller numbers (6 responses) perceived either negative impacts or positive impacts (5). Overall, instructors expressed ambivalence, with no clear consensus on whether generative AI improves or undermines project quality.

Institutional policies on the use of generative AI were reported as uneven. While 12 instructors indicated that their institutions had clearly stated policies, 9 described policies as somewhat unclear, and only one reported the absence of any policy. This suggests that although guidance exists in many contexts, its clarity and consistency vary significantly.

In terms of assessment practices, about half of the instructors had begun adjusting their rubrics or methods in response to generative AI. Nine reported slight adjustments, and three significant adjustments, while others indicated no changes—seven with no plans to adjust, and three planning to do so in the future. These responses suggest an uneven but emerging adaptation to account for generative AI in assessment design.

Finally, most instructors reported that they do not currently check for AI-generated content in student submissions (13). Among those who do, six rely on manual checks, such as identifying style inconsistencies, while three allow generative AI use, provided that it is disclosed by students. Together, these findings highlight that although generative AI is present in capstone contexts, its impact on assessment practices is still evolving. Instructors remain divided in their perceptions of its influence on quality, and institutional and pedagogical responses are still developing.

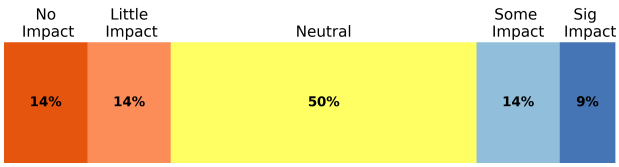


Figure 5: GenAI effect on capstone project quality

The distribution of responses indicates mixed perceptions. Half of the participants selected the neutral option, suggesting uncertainty or ambivalence about the role of generative AI. At the same time, a considerable proportion expressed skepticism, with 14% disagreeing and another 14% strongly disagreeing that generative AI improves quality. By contrast, 14% agreed and 9% strongly agreed, indicating that a minority view generative AI as positively impacting capstone project quality. Overall, these findings reflect a wide range of perspectives, with responses clustered at the neutral midpoint, underscoring the contested nature of generative AI's influence on student project outcomes.

6.1.8 Proposed Strategies to Address Challenges. When asked which strategies could most effectively address the challenges of assessing team-based projects, participants identified several key themes (Table 7). The most frequently endorsed strategy was integrating self- and peer-assessment components with proper calibration (14 responses), suggesting that many instructors see value in structured mechanisms for capturing individual contributions within teams. Incorporating regular formative assessments and iterative feedback sessions (12) and leveraging digital tools to track individual contributions (11) were also viewed as important, reflecting a desire for more continuous and data-driven insights into team dynamics. Other strategies included providing training sessions for students and instructors on AI ethics and usage (10), developing detailed and flexible rubrics (8), and allocating additional resources such as grading or design support (7). Together, these responses indicate a preference for multifaceted solutions that combine calibrated peer input, ongoing formative feedback, technological support, and enhanced instructional resources to strengthen the assessment of team-based projects.

Table 7: Suggested Improvements for Capstone Assessment

Suggestion	Frequency
Integrating self- and peer-assessment components with proper calibration	14
Incorporating regular formative assessments and iterative feedback sessions	12
Leveraging digital tools to track individual contributions within teams	11
Training sessions for students and instructors on AI ethics and usage	10
Developing and implementing more detailed, dynamic rubrics	8
Allocation of more resources (grading support, design support, etc.)	7

6.1.9 Critical Factors in Implementing New Assessment Strategies. Figure 6 presents responses to the question, *'What are the most critical factors to consider when implementing new assessment strategies in your courses?'*. The two most frequently cited considerations were time and resource availability and alignment with learning outcomes, each selected by 17 participants. Scalability was also seen as important, highlighted by 12 respondents. Fewer participants pointed to transparency (8) and student acceptance and training (7), though these were still noted as relevant. Overall, the findings suggest that while logistical constraints and alignment with curricular goals are viewed as paramount, issues of scalability, clarity, and student buy-in also play a role in shaping the successful adoption of new assessment strategies.

6.1.10 Insights from the Quantitative Analysis. Having presented the results of our quantitative data collection from 22 instructors, we now summarize the insights gained from their survey responses. The responses highlight a consistent reliance on instructor evaluation as the dominant basis for grading in capstone courses, complemented by a mix of formative and summative approaches and a variety of deliverables. While some alternative methods, such as peer, self-, and stakeholder assessments, are present, their role appears more limited and contested. To better understand the reasons behind these patterns and to capture richer insights into the challenges and practices of assessment, we next turn to the qualitative findings from instructor interviews.

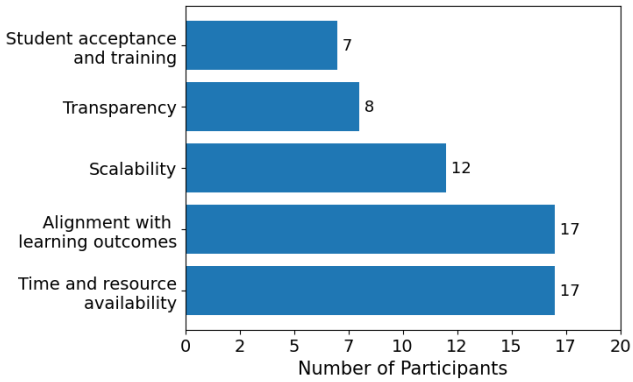


Figure 6: Critical factors to consider when implementing new assessment strategies

6.2 Instructors' Qualitative Responses

We initially conducted interviews with 14 instructors, but ultimately based our analysis on 13 of them, as shown in Table 8. One instructor (INST5) was excluded because she declined to answer many of the assessment-related questions, and her interview was therefore limited in scope. She also expressed concerns about sharing her practices due to potential conflicts with her own plans for publication. To maintain the integrity of our coding and identifiers, we retained the original numbering, leaving INST5 omitted while continuing the sequence up to INST14.

Table 8: Demographics of instructor interviewees

Participant ID	Gender	Continent	Years Teaching Capstone
INST1	Male	Europe	1 - 3 years
INST2	Male	Europe	1 - 3 years
INST3	Male	Europe	1 - 3 years
INST4	Male	Europe	1 - 3 years
INST6	Male	North America	7-10 years
INST7	Female	Europe	1-3 years
INST8	Male	Europe	1-3 years
INST9	Female	Europe	1-4 years
INST10	Female	Europe	20 years
INST11	Male	Europe	2 years
INST12	Male	Europe	5 years
INST13	Male	Europe	16 years
INST14	Male	Europe	6 years

The thematic analyses of the fourteen in-depth interviews with instructors resulted in four main themes:

Instructor Background: The professional background of the instructors varied quite vastly, not only with respect to their

experience with running and assessing capstone project courses, but also with respect to the roles assumed. Moreover, participants varied with respect to how much they interacted with external stakeholders.

Assessment Methods: Core to the interviews was the set of statements related to what was assessed and how. Participants discussed not only their own choices and preferences but also the extent to which external stakeholders influenced assessment practices.

Instructor Challenges: Reflecting on challenges in their capstone projects, instructors commented on team dynamics, peer evaluation, and the need to balance individual and team contributions. They also noted issues with non-contributing students.

Evolution of the Assessment methods: The statements in this theme were reflections on how assessment methods and rubrics used for them had evolved over multiple course offerings. This theme differed from the above theme "Assessment Methods" in that the statements were reflective and not descriptive.

More details regarding the different facets of statements assigned to these themes will be discussed in the following subsections. We provide illustrative quotes from individual interviews for some topics, to give anecdotal examples of the issues discussed.

6.2.1 *Instructor Background.*

Instructor Experience. Instructors in our interview set possessed a wide range of experiences in teaching and assessing capstone project courses, with involvement ranging from as little as two years to over two decades. Some instructors said that they oversee more than 50 projects annually and also manage large teams of supervisors, while others have been in capstone project courses in different institutions across different countries.

Roles and Responsibilities. Capstone instructors serve in multiple capacities – as coordinators, supervisors, mentors, and examiners, balancing course-level oversight with team-level support. At the course level, they manage timelines, team formation, stakeholder engagement, deliverable reviews, and the consistency and fairness of assessment. At the team level, they guide projects through regular meetings and feedback, monitor team dynamics, address conflicts, and evaluate both technical outputs and soft skills such as collaboration, professionalism, and communication.

Engagement with External Stakeholders. Capstone courses frequently involve external stakeholders as project clients, who propose real-world projects and provide feedback to students. For example, one instructor, INST4, reported collaboration with 46 organizations and over 230 industry partners over the past five years.

External stakeholders typically present project proposals, offer regular feedback on project progress, and comment on final demonstrations. Communication between students and stakeholders varies: in some courses, instructors serve as the main liaison, while in others, students interact directly with the stakeholders, who act as clients for the projects. Some courses host public exhibitions for external stakeholder-student interaction. In some courses, these were paused during the pandemic and have not resumed.

6.2.2 *Assessment Methods.*

What is assessed? Interviewees identified a range of assessed artifacts, including sprint logs, written reports (mid-term or final), video demonstrations, oral presentations, and code repositories. Beyond these outputs, instructors emphasized evaluating both technical skills, such as implementation, testing, and documentation, and non-technical skills, including teamwork, communication, reliability, and professionalism.

This balance reflects the importance of not only what students produce but also how they collaborate and behave in a professional context, with teamwork and professionalism consistently highlighted as central assessment criteria.

> *It's both the technical deliverables and the soft skills, i.e., communication, teamwork, reliability. The final grade reflects their professional conduct as much as their code or designs.* [INST12]

> *Learning outcomes include behaving in a professional manner, contributing to teamwork, and delivering the committed functionality. Both technical contributions and behavioral aspects are considered.* [INST11]

Who assesses? The main instructor of the course holds primary responsibility for assessment. Supervisors, who meet with student teams throughout the course, also play a significant role in evaluating student work. Stakeholders are sometimes involved in assessment, typically providing structured or unstructured feedback rather than formal grading.

> *Primary assessors are course instructors/mentors. External clients provide feedback but do not directly assign grades. Peer evaluations are used as formative input but not for grading.* [INST13]

In some cases, a third person, someone not involved with the course during the semester, acts as an assessor to ensure fairness and facilitate open communication between the supervisor and student teams. In such cases, the supervisor does not participate directly in the assessment. Instead, external assessors and the project coordinator assign the final grade after discussions.

> *Examiner assesses final grade... Mentors keep roll call and logs, read weekly reports, and identify issues. Customers can provide input, but usually have limited individual insight...* [INST11]

Grade Distribution. Grade distribution was generally divided into two main categories – *individual versus team* grade distribution, and *product versus process* grade distribution. In most cases, product-focused grading resulted in a team mark, while individual grades, when assigned, were typically based on individual reflection reports or assessed competencies.

> *It is 40% (product), 40% (process), 20% (reflection), and I think it is fair ... They need to focus on both process and product. And then they need to have a good percentage of thinking about both in the reflection phase.* [INST9]

In some courses, all team members received the same grade, with no individual assessment, often due to challenges in identifying individual contributions or the complexity of implementing a strategy to assign individual grades.

External Stakeholder Influence on Grades. Instructor responses revealed varying levels of stakeholder involvement in grading. Some instructors excluded stakeholders entirely, arguing that their lack of teaching background and the diversity of projects made fair grading difficult. In other cases, stakeholders provided structured or informal feedback that informed instructors' assessment but did not directly contribute to grades.

> *Stakeholder involvement in grading is dangerous: no teaching background, different projects need different assessments, and fairness is not ensured.* [INST10]

In some programs, university policy explicitly prohibited stakeholder involvement in grading.

> *They can give their views, but university regulations wouldn't allow externals to set marks.* [INST4]

Use of Rubrics. The use of rubrics varied significantly across interviewees. Some found them difficult to implement and preferred to keep them abstract, while others did not use them at all. Overall, there was a shared sense that some form of guideline is helpful, but the extent and specificity of rubric use differed widely. This variation reflects the underlying heterogeneity of capstone courses

> *So there are some common criteria that you can apply to any project, but then if you go more into the specifics of a software project... it can be very different from one project to the other, so, some criteria might be very important for one type of project... and may not even be relevant for another project ... depending on what the customer proposes and what type of work is required from the group. So the ... rubric/evaluation criteria is abstract. They are more like high-level in a way that they can cover the common aspects of this project.* [INST9]

Interviewees also discussed who designed or 'owned' the rubrics. In some cases, the rubric was inherited from earlier iterations of the course and gradually modified, while in others it had been fully redesigned by the lead instructor with input from departmental colleagues. In a few cases, the rubric was reshaped or restructured on an annual basis by incorporating feedback from supervisors and staff involved in assessment. As one instructor explained,

> *What I did is actually a study of all the feedback that came in the years before me, and tried to integrate some of that. So it captures the different aspects that different people are looking for.* [INST4]

Main Takeaways (Assessment Methods). Assessment in capstone projects is multifaceted, including the assessment of code, reports, logs, and oral presentations. While the formal responsibility for the assessment lies with with main instructor who may delegate this to neutral third parties, external stakeholders can provide feedback and contribute to the final grading. There is neither a one-size-fits-all method for determining a grade based on both individual contributions and team performance nor an agreement on whether or to which extent rely on rubrics.

6.2.3 Instructor Challenges. Our interviews revealed key challenges in assessing capstone projects, including balancing individual and group contributions, preventing unfair peer targeting, recognizing both leadership and supporting roles, calibrating self- and peer evaluations to avoid misuse, and managing fatigue from overly frequent assessments.

Balancing Individual vs team contributions. Interviewees reported multiple strategies to balance individual and team contributions within project-based work. A prominent approach was using agile methods, including short sprint cycles, regular stand-ups or progress meetings, and version control tracking. For example, story point graphs and commit histories allowed instructors to monitor engagement without excessive intervention.

> *I do weekly check and meetings with all the teams and over time it becomes clear to me if someone is not doing enough or too much.* [INST6]

Instructors promoted equitable participation by encouraging teams to self-organize, rotate leadership, and match tasks to skills. In cases of imbalance, instructors could adjust individual grades, with decisions supported by documentation or team feedback.

Instructors also emphasized balancing autonomy with oversight. While student ownership was seen as vital, safeguards such as early warning systems, milestone reviews, and transparent grading policies were used to prevent unfairness and disengagement.

> *A student could do quite poorly in terms of their individual contribution, but could be greatly uplifted if the team performed very well as a whole, or vice versa. It was really important that the supervisors were able to look at those individual contribution marks and nudge them upwards or downwards, depending on their observations of individual performance.* [INST3]

> *It's not just assessment, it's not just scaling of marks based on contribution. It's also what you do proactively to make sure that those teams work well and function well.* [INST4]

Impact of assessment on team dynamics. Assessment was reported to shape team dynamics in both positive and negative ways. Grading based on contributions was described as a positive influence, promoting accountability, discouraging passive behavior, and encouraging timely engagement.

> *The same grade policy enforces students to stay silent even in case of uneven workload or any conflict.* [INST7]

Formative practices such as interim assessments, self-reflection activities, and peer feedback opportunities were also seen to support introspection and behavioral adjustment among students. On the other hand, some assessment approaches generated tension in teams. For instance, peer reporting of non-contribution sometimes led to emotional fallout and fractured relationships.

> *When peer feedback is too blunt, it can create tension — some students stop communicating. We're trying to make feedback more constructive to avoid breaking trust.* [INST12]

Reported students often felt targeted or unfairly judged, especially when peer feedback lacked mentoring support. Instructors noted that assessment structures can intensify team conflicts, particularly when coupled with tight deadlines, unclear task dependencies, or ambiguous role expectations.

> *Report-based assessment leads to poor team dynamics... reported students feel betrayed or attacked. Negative consequences for cohesion – students may withdraw or become cold.* [INST11]

While some instructors believed that grading schemes moderately affected team dynamics, others observed more direct correlations. The presence or absence of individual grading, the visibility of contributions, and the degree of instructor involvement all shaped group behavior. Finally, team dynamics were influenced by both formal structures and informal social processes, including peer pressure, trust, and communication styles.

> *Although the general idea is for the whole group to get the same grade for the task, I reserve the right of changing this depending on the group, the individual efforts contributing to that assignment.* [INST6]

Interpreting Peer Evaluation. Peer evaluation emerged as a practice used with caution and typically served as a formative tool rather than a grading mechanism. In courses where peer evaluation was implemented, students reviewed their teammates' work using predefined criteria and were asked to offer constructive feedback. This feedback process often continued periodically for the entire course, starting early in the project and continuing to the final submission. Students were also encouraged to reflect on the feedback they received. This could be done by developing actionable goals or flagging perceived bias or inaccurate input. Instructors drew on these evaluations to inform their judgments but rarely incorporated them directly into grading.

> *We've tested peer evaluation forms, they're anonymous and focus on specific behaviors, not personalities. Right now, they inform my judgment but aren't used as a direct grading factor.* [INST12]

Supervisors often triangulated peer input with additional evidence, such as version control histories, communication logs, or mentor observations, to validate claims about individual contributions. Several instructors emphasized that peer assessments sometimes reflected underlying interpersonal tensions within teams, making it essential for educators to interpret such data within the broader context of team dynamics.

> *[It] creates some complex social dynamics because I think the students sometimes believe or perceive that they are marking each other. This is not accurate in the sense that the quote-on-quote scores from that peer review process only inform the individual contribution mark and the supervisor's academic judgment.* [INST3]

Despite these efforts, the reliability and fairness of peer evaluation remained a concern. Many instructors noted that limited visibility into other teammates' work could affect peer review data, leading to over- and underestimating contributions. Moreover, the quality of peer feedback varied significantly. Students sometimes wrote vague or overly critical comments, while others flooded the system with dense feedback, potentially skewing perceptions. A recurring challenge was the "tit-for-tat" dynamic, where students reciprocated low or high ratings based on personal alliances rather than actual performance.

> *They would overestimate or underestimate how the person had performed. They might leave the scoring process till the last minute, and as a result, give very shallow or sparse feedback, or some of them might take it super seriously and might give a ton of feedback.* [INST3]

In some institutions, 360-degree evaluation models were tested, combining peer feedback with supervisor judgments to assign individual marks. However, the consensus remained that peer evaluation, when used summatively, should be approached carefully due to its susceptibility to social dynamics and uneven interpretation. As a result, many instructors chose to keep peer assessment as a non-graded component or used it sparingly, e.g., for mid-semester feedback only.

In some cases, instructors only experimented with peer assessment to discontinue it due to concerns about fairness, student dissatisfaction, and the administrative burden it placed on students and educators. In other implementations, peer evaluation was available as an optional diagnostic tool or reserved for instructor review only, without student-to-student scoring.

> *Because of the nature of the peer assessment, we don't want students kind of brigading on an individual, therefore it is used by the supervisor to determine the grade based on their judgment, which includes reviewing things like version control and engagement with the team communication platforms .* [INST1]

While peer evaluation was seen as valuable for supporting team accountability and identifying participation issues, it was not widely adopted as a core assessment mechanism. Instead, it was more commonly positioned as a supplementary input, best used to trigger mentor intervention or guide reflective dialogue rather than to determine grades directly.

Addressing student concerns on fairness. Instructors expressed concerns about managing students' perceptions of fairness in capstone assessments. They admitted that group grading often masks differences in individual contributions, leading to dissatisfaction among high-performing students.

> *The formal process of individual contribution marks that ended up being pretty crucial to fairness. I think if we hadn't had that, we probably would have had significant dissatisfaction from students about how we did things.* [INST3]

To address this, they used tools like peer feedback, version control logs, and work-hour reports, and held mid-project reviews and team meetings to assess group and individual contribution. Instructors (e.g., INST 4) emphasized that these trace-based indicators were never used in isolation; instead, they were interpreted together with supervisors' qualitative observations, individual contribution reports, and team communications maintained throughout the project. It was described as essential because no single source of evidence was seen as sufficient for capturing the full range of student contributions. In complex cases, instructors often call meetings with the

students involved to mediate conflict, clarify expectations, discuss strategies for more effective collaboration, encourage open communication within the team, reminding them of professional conduct guidelines.

> *It is not just the peer review, it's the supervisor opinions, Git repositories, individual report, meeting minutes that they maintain from different meetings that they have throughout the year. There is a lot of stuff that sit behind it.* [INST4]

Despite these efforts, some students stayed silent about unfair workloads, and conflicting team narratives complicated instructor decisions. Institutional constraints sometimes led instructors to pass low-contributing students to avoid appeals. Ultimately, fairness was less a strict standard and more a negotiated balance of evidence, team dynamics, and instructor judgment.

Handling non-contributing students. One of the major challenges instructors face is identifying and addressing non-contributing students in a way that does not disrupt group dynamics. Instructors emphasized the importance of early identification, as patterns of contribution often shift over the course of a project – some students may become more engaged later, while others reduce their involvement.

> *Mentors warn them first, then I have a formal conversation...If it continues, they can fail before the course ends.* [INST12]

To detect non-contribution, instructors employ multiple strategies, such as monitoring weekly hours reported by students, reviewing Git logs and story point graphs, and analyzing peer evaluations. However, instructors also emphasized that these indicators needed to be interpreted in context rather than taken at face value. As INST1 noted earlier, supervisors routinely combine repository activity with observations of students' communication patterns and team interactions, to form a more complete picture of participation. Weekly meetings were also described as a crucial source of insight by INST6, with instructors reporting that over time it becomes evident who is engaging and who is struggling.

Students are also encouraged to report instances of unfair workload distribution, with instructors relying on their honesty to surface issues. Multiple perspectives are often triangulated, drawing on the views of external clients, faculty mentors, and peers to form a more complete picture of each student's involvement. Alongside monitoring, many instructors promote empathy-based approaches that encourage teams to check in on struggling peers and implement paired support strategies to assist less-experienced or vulnerable students. When necessary, conflicts are mediated directly, and persistent non-contribution may result in lower individual grades. These practices reflect a balance between maintaining fairness, preserving team cohesion, and supporting student growth.

> *I always encourage team members to ask whether the [non-contributing] person is okay. By simply asking that question that can make a huge difference...Occasionally there is indeed circumstances beyond their control that they cannot do anything about that influences their participation.* [INST4]

Main Takeaways (Instructor Challenges). Instructors noted challenges with providing motivation and fair feedback to participants when individual contributions are hard to separate from team performance. In general, a "same grade for all" policy was considered to be less suitable than detailed, individual feedback. At times, peer evaluation was found to be too direct, potentially unreliable, and often negatively impacting team cohesion. As a consequence, managing non-contributing students was considered to be the responsibility of the instructor but identifying such students was found to be challenging. Instructors also emphasized that no single source of evidence, e.g., logs or reported hours, was sufficient on its own, making multiple inputs essential for evaluating individual contribution.

6.2.4 *Evolution of the Assessment Methods.*

Rubric Evolution. In most cases, instructors reported that their rubrics have not changed substantially over time. If the rubrics do evolve, changes are based on the instructor's reflections and student feedback.

> *We're slowly building a proper rubric with clearer contribution criteria. We want to define what counts as 'meaningful' work in a way that's transparent.* [INST12]

> *I was always editing them [rubric criteria], changing them based on observations and feedback from staffs and students.* [INST3]

Sometimes changes are introduced to include more structure to the assessment process, as well as changes in the percentage of marks.

> *Currently moving towards more contribution analysis... want to define 'meaningful contribution' and make it verifiable.* [INST11]

Other changes in assessment strategies. Several shifts have been introduced over time to adapt to evolving contexts. These include altering the weightings of marks, developing more structured marking schemes, and incorporating interim peer assessments as proactive measures to identify contribution issues earlier. Instructors also reported the necessity of cross-referencing individual and group reports with peer evaluations, as well as moderating across all deliverables to ensure fairness and consistency.

> *We have also put more focus on code quality. Because at a certain point, we experience that the students were putting too much focus on the report rather than on what they actually developed.* [INST10]

External pressures have also shaped assessment strategies in capstone courses. Instructors identified that the rapid evolution of technology requires updates to assessment design, while the COVID-19 pandemic forced many institutions to move final showcases online, creating new challenges for evaluating both technical artifacts and presentation skills.

> *I am keeping the same major components, but some assignments have been changed. It's probably connected to the technologies some groups are using. The agile field also changes, so I get to change some of the requirements of the projects.* [INST6]

Main Takeaways (Evolution of the Assessment Methods). Instructors were found to strive to improve the fairness, transparency, and reliability of their assessment methods to address the students'

concerns and expectations. Instructors emphasized, however, that such changes may also bring unintended consequences.

7 Students' Perspectives

This section presents insights from students gathered through both the online survey and in-depth interviews. As with the instructor data, we report our findings separately: the survey results provide breadth by capturing perspectives from a larger group of students, while the interviews offer depth by highlighting individual experiences. The qualitative findings are further organized around themes that emerged from the thematic analysis of the interview data. Together, these complementary perspectives provide a holistic view of how students experience and interpret capstone assessments.

As with the instructor perspectives, we provide illustrative quotes from individual interviews.

7.1 Students' Quantitative Responses

7.1.1 Students and Course Background. An online survey was used to collect responses from 44 students. Most reported taking capstone courses at the bachelor's level (34), with some at the master's level (9) and only one at the associate level. Team sizes varied, though the most common arrangement was six members (17 students), followed by teams larger than six (10). Smaller teams of 3–5 were less frequent. When asked about team formation, the majority indicated their teams were self-chosen (31), while fewer noted instructor-assigned teams (6). A small number described other approaches, including random assignment (1), forming groups with prior acquaintances (1), or selecting teammates based on shared project interests (1).

7.1.2 Assessment Components. Assessment components are those aspects of the course that relate to measuring attainment and providing feedback. Students were asked to identify the types of deliverables required to assess their capstone courses. As shown in Table 9, nearly all respondents reported oral presentations (98%), completed project builds (93%), written reports (86%), and project demonstrations (84%) as common deliverables. Reflection essays or journals (64%) and progress logs (57%) were also frequently mentioned, though less consistently required. A small number of students (7%) noted other deliverables such as peer feedback or peer reviews. These results suggest that while traditional deliverables such as reports, presentations, and project outputs remain central, reflective and process-oriented components are also widely incorporated.

We also asked about the extent to which different assessment types contributed to determining students' final grades. Figure 7 presents the students' perceptions. The results show a strong emphasis on instructor evaluation, which students consistently identified as having the most significant impact on their final grades. In contrast, peer evaluation and self-assessment were viewed as contributing only moderately, while external stakeholder feedback was perceived as the least influential. These findings suggest alignment between instructors and students in recognizing the central role of instructor evaluation, alongside more limited contributions from peer-, self-, and stakeholder-based assessments.

When asked whether the assessment criteria for the capstone project were clearly communicated and understood, most students agreed. Twenty-one students strongly agreed, 16 agreed, while only four chose neutral, and three strongly disagreed.

Table 9: Student-reported deliverables in capstone courses

Deliverable	Frequency (%)
Oral presentations	43 (97.7%)
Completed project build/deliverable	41 (93.2%)
Written reports (e.g., documentation, technical report)	38 (86.4%)
Project demonstrations (e.g., live demo, walkthrough)	37 (84.1%)
Reflection essays or journals	28 (63.6%)
Progress meeting minutes or team logs	25 (56.8%)
Other (e.g., peer feedback, peer reviews)	3 (6.8%)

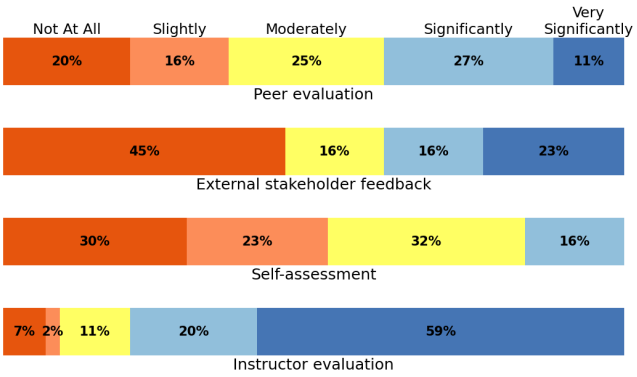


Figure 7: Students' opinion on the degree by which an assessment type contributed to determining students' final grades. Percentages may be 100% ± 1 due to rounding.

7.1.3 Satisfaction and Fairness. Students generally expressed high levels of satisfaction with the feedback they received on their individual contributions to the capstone project. Twenty-one students reported being very satisfied and sixteen indicated they were satisfied. Only a small minority expressed neutrality (4) or dissatisfaction (3). When asked whether the assessment process adequately reflected their individual contributions, most students responded affirmatively, with 33 answering "Yes" and 8 "Somewhat," while only 3 felt their contributions were not adequately captured.

Students also reported strong confidence in the fairness of instructor evaluations for judging overall team performance. Most described instructor evaluation as fair or very fair, while only a small minority expressed neutral or negative views. In addition, students emphasized the importance of rubrics with clearly defined criteria in evaluating capstone projects. The majority considered rubrics to be very important, with only a few remaining neutral or regarding them as less useful. Taken together, these results suggest that students are largely satisfied with the fairness and clarity of assessment practices, and place strong value on explicit rubrics to guide evaluation

Students were also asked about the perceived fairness of different methods for judging individual contributions in capstone projects

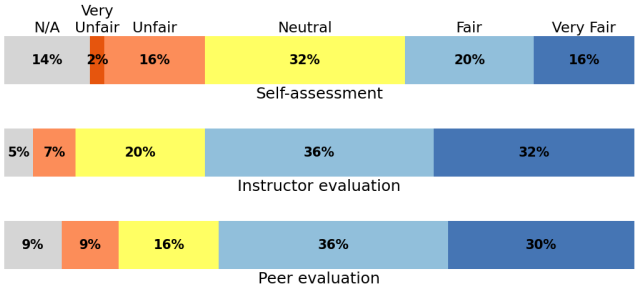


Figure 8: Student perceptions of fairness for different methods of evaluating individual contributions. Percentages may be 100% ± 1 due to rounding.

(Figure 8). Instructor evaluation was rated most favorably, with 36% viewing it as fair and 32% as very fair, though 20% considered it unfair and smaller proportions selected very unfair (7%) or N/A (5%). Peer evaluation also received generally positive ratings, with 36% describing it as fair and 30% as very fair, while 16% considered it unfair and 9% very unfair. In contrast, self-assessment was viewed with greater skepticism: although 20% rated it fair and 16% very fair, nearly one-third (32%) responded neutrally, and a combined 18% judged it unfair or very unfair. Together, these findings suggest that students tend to view instructor and peer evaluations as relatively fair mechanisms for assessing individual contributions, while self-assessment is regarded more cautiously and often with neutrality.

7.1.4 Use of Generative AI Tools . When asked about the incidental use of generative AI tools during capstone projects, many students indicated that such tools, particularly ChatGPT, were used at least occasionally, while others reported no use or were uncertain. A small group noted that their projects were explicitly focused on generative AI, making the question not applicable. Among those who had used these tools, disclosure practices varied: some students were transparent with both instructors and teammates, while others chose to share only within their team. However, a substantial portion of students reported not using AI tools at all, making disclosure irrelevant in their case.

Students' responses also revealed uncertainty around institutional expectations. While some reported that clear guidelines on incidental AI use were in place, others described the policies as vague or were unsure of what rules applied. Taken together, these findings suggest that while generative AI is beginning to play a role in capstone projects, its use remains uneven, and students perceive a lack of clarity in institutional guidance on when and how such tools can be appropriately used. These student perspectives align with instructors' own ambivalence reported in our survey and interviews, where many expressed uncertainty about appropriate boundaries for AI use and emphasized the need for clearer policies and shared norms.

Students were asked whether they agreed with the statement that the use of generative AI tools compromises the fairness and integrity of capstone assessments as shown in Figure 9. Responses were mixed, reflecting a diversity of perspectives. A significant share of students disagreed, indicating skepticism about whether AI use inherently undermines fairness, while about a third adopted a neutral stance, suggesting uncertainty or ambivalence. At the same time, more than a third of respondents agreed or strongly agreed that generative AI poses risks to fairness and integrity in assessment. These results highlight the contested nature of generative AI in the capstone context, with students divided between concern, dismissal, and uncertainty about its impact

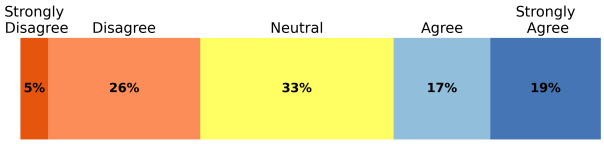


Figure 9: GenAI effect on fairness and integrity of capstone assessment.

7.1.5 Suggestions for Improving Capstone Assessment. When asked which strategies could most effectively improve the evaluation of team and individual contributions in capstone projects, students highlighted several recurring themes (Table 10). The most frequently endorsed approaches were incorporating regular formative feedback sessions (68%), clearer communication of assessment criteria and expectations (57%), and the use of digital tools to track individual contributions (55%). More than half also identified the development of more detailed rubrics and guidelines (52%) as an important improvement. Strategies mentioned by smaller but still notable proportions included incorporating self-assessments and reflective practices (32%) and enhancing peer evaluation with calibration sessions (32%). A few students suggested other approaches, such as clearer communication of stakeholder expectations, more explicit success criteria, and more frequent in-class work sessions. Taken together, these results suggest that students favor strategies that enhance transparency, provide ongoing feedback, and leverage tools to capture individual effort more effectively.

Table 10: Student-identified strategies to improve evaluation

Strategy	Frequency (%)
Regular formative feedback sessions	30 (68.2%)
Clearer communication of assessment criteria and expectations	25 (56.8%)
Use of digital tools to track individual contributions	24 (54.5%)
More detailed rubrics and guidelines	23 (52.3%)
Incorporation of self-assessments and reflective practices	14 (31.8%)
Enhanced peer evaluation processes with calibration sessions	14 (31.8%)
Other (e.g., stakeholder expectations, success criteria, more in-class work sessions)	3 (6.8%)

7.2 Students' Qualitative Responses

Twelve students participated in the in-depth interviews (Table 11). The analysis of interview transcripts was conducted thematically, and in the following sections, we present the findings organized around the key themes that emerged from this process. Students recalled and described their understanding of the capstone practices they experienced in their course. This typically focused on the work they were doing for assessment and the structure of the course in terms of activities and their interaction with supervisors. They also reported on their perceptions of fairness when assessing individual contributions.

Table 11: Demographics of student interview participants

Participant ID	Gender	Continent
STD1	Male	Europe
STD2	Male	Europe
STD3	Male	Europe
STD4	NR	Europe
STD5	NR	Europe
STD6	NR	Europe
STD7	NR	Europe
STD8	NR	Europe
STD9	Male	Oceania
STD10	Female	North America
STD11	Female	Oceania
STD12	Male	Oceania

The thematic analysis of the twelve in-depth interviews with students resulted in three main themes:

Reflective Insights: Statements that were assigned to this theme presented themselves as descriptive considerations of individual and group experiences as well as of aspects of assessment and supervision.

Suggested Improvements: Statements that were assigned to this theme usually, but not always, were elicited by the interviewers' prompt to reflect upon what changes in the course setting might help improving the educational experience. Rather naturally, statements assigned to this theme pertained to the course structure in general as well as to specific aspects of supervision, assessment, and feedback. At times, such statements also were found as part of earlier parts of the interviews, i.e., as afterthoughts to the descriptive considerations mentioned above.

Advice for Incoming Students: Again, statements assigned to this theme were found naturally in response to the corresponding prompt. These statements covered strategic and practical recommendations but, interestingly enough, also address helpful mindsets for navigating the capstone course.

More details regarding the different facets of statements assigned to these themes will be discussed in the following subsections.

7.2.1 Reflective Insights. Students were deeply introspective as they reflected on their capstone experiences. We initially considered various sub-themes related to sentiment, such as challenges and commendations, or to the matters arising, such as transparency, fairness, and so on. Still, we decided to use a structure based on different aspects of the course structure (e.g., assessment, feedback, supervision, etc.) for ease of locating where and what students were discussing.

Engagement with Stakeholders. The projects that the students described varied considerably in terms of their engagement with stakeholders and whether or not they had set requirements. Most students reported having an internal stakeholder, an external industry stakeholder, or a target market, with corresponding foci on what they had been set to do. Typically, students working with internal stakeholders worked on research-related projects, but a few students reported having no stakeholder and instead worked on a project of their choosing.

Available Resources and Tools. Students typically discussed resources in terms of what their supervisor was able to do or secure for them and the toolsets they typically used to execute their project. These included things like task boards, communication apps, social media, versioning, and continuous deployment systems, development environments, engines, or other frameworks. These tended to vary across the courses and institutions in which the students were based.

Assessment. Student reflections of assessment can broadly be categorized around three themes. The first was *transparency*, particularly relating to the clarity of the assessed task and the assessment criteria. The second concerned differences in *perceptions* of the process, expressed with both positive and negative sentiment. Finally, students highlighted factors they felt *influenced* their grades—such as assessor bias and the extent to which individual contributions were accurately *recognized*.

Assessment: Transparency. Students expressed the importance of clarity in the way information is conveyed to them, alongside a desire to understand how they will be assessed. Some of the students expressed concern about their comprehension of assessment goals, processes, and outcomes, with the potential implications being undeserved grades. Rubrics were addressed in all interviews, be it for their perceived benefits or for the stress induced by uncertainties about grading processes for which no rubrics were present:

> *So we knew from the start exactly how we would be assessed.*[STD6]

> *So there were a lot of components in there that we just didn't know and so the challenge for me was, I just wanted to get a good grade and I just didn't know what I had to do, I just did as much as I could.* [STD4]

Assessment: Perception. Perception of assessment varies, with many acknowledging varying degrees of fairness and varying sentiment towards the overall process. Timeliness of feedback is a particularly prominent sub-theme, as were various attitudes towards peer evaluation—including the reluctance of many students to give critical feedback or mark their peers. This was the case both in settings in which peer review was a mandatory component of grading and in settings in which assessors offered the students the opportunity to volunteer anonymous peer feedback:

> *And so I feel that the peer review contributed heavily to it. Whilst this wasn't a problem in my team, I worry that it may be exploited elsewhere.* [STD1]

> *I explicitly didn't use this [opportunity to give anonymous feedback on my peers' performance] because I would have found it very unsolidary (sic) even if someone had stood out negatively and somehow a bit too much applied game theory for my liking.*[STD6]

Assessment: Influences. Various aspects of assessor consistency and impartiality were explored, particularly with respect to the different perspectives they could offer. A few students mentioned that matchmaking was important to ensure appropriate expertise was being used throughout the assessment process. There was recognition of the time it takes to assess work rigorously, and that shorter times might undermine the assessment.

> *I think the demo day, however, was a bit short. Because the people playtesting it were only played for 15 minutes and the game was about 45.* [STD1]

In other cases, students reported that assessors were not present regularly and that may have skewed the assessment:

> *I worry that the supervisor doesn't have I guess a full image of the actual going ons in the project.* [STD1]

> *[They] weren't really there to see how much everyone took part in the joint discussions and ideas that went into the project. And I think that led to a few misunderstandings and some unfair assessments.* [STD5]

Assessment: Recognizing Contributions. Students were concerned about how intangible contributions are recognized, particularly work like managing version control, continuous integration, leadership, project management, and so on.

> *Actually getting it all to work. That was mainly me. Who did that. And I felt like the actual integration process and the importance, that that was maybe overlooked.* [STD2]

A recurring theme was the importance of not to overlook "hidden contributions" that may not have explicit correspondences in the grading scheme:

> *At the very beginning, it got lost a bit that I did a lot of organizing and that's just not included in the evaluation process, that you need a person who is a bit of a mom, so to speak, who organizes everything.* [STD7]

Some students advocate for tools as a lens to understand individual contributions to a project. It is often challenging to assess the wide range of roles that students might undertake during a project, and some suspect that it is challenging to reliably quantify their contributions:

> *Did they do badly? Do they good? It's very quantitative, very. Like, give me a number that's representative of something when in reality.* [STD3]

Development Process. Students reflected considerably on their development process, with two key areas emerging as particular sub-themes, *team productivity* and *navigating the openness of the capstone projects*. One student also mentioned the tension in managing stakeholder needs. Various topics emerged, including criticisms of the agile approach to software development, the need to forge one's own path, learning by doing, enjoying the development work/process, and managing scope and risks.

Feedback. Students mentioned the respective ease of understanding their feedback, the frequency and volume of their feedback (and in particular formative feedback), the importance of explaining feedback and providing reassurance, as well as proactively engaging with staff and stakeholders to get feedback.

Notably, students explicitly sought feedback coupled with opportunities for improvement and were concerned about (peer) feedback not being critical enough:

> *Yes, especially with the oral grade, because it came up once in between, I knew that if I wanted a better grade, I would have to make more of an effort.* [STD7]

> *I think those methods are overall good, but I do notice a trend where there were a lot of team members not really being critical enough of each other, so they'd pretty much give everyone high grades.* [STD1]

Preparedness. Students mentioned the expectation of the capstone, including its novelty and the sense of taking responsibility for a project. They also related this to their prior experience, which was often minor or insignificant. One participant noted that their internship was quite different to their capstone:

> *I would say the closest that I ever did was the software internship, which didn't even begin to address the requirements of the capstone project.* [STD5]

Group Dynamic. Students often commented on the dynamic of the group they were a member of. This frequently reflected the role they assumed, whether as a programmer, designer, project manager, or someone who led the project as a vision holder. The diversity and fluidity of these roles are something they surfaced, but often highlighted pressures to adapt due to the evolving dynamic of the group, especially with respect to disengagement and free-riding.

> *I've got three people who can't do anything and they're going to mess up my grade or something.* [STD4]

Trust and conflict were also mentioned, with the influence of pre-existing friendships and friendship groups complicating interpersonal exchanges.

> *With team dynamic, it can sometimes be hard when you are in a project with friends. [...] Because of how the structure is of it being an undergraduate degree and not quite the same as a job [...] so significant people have different ideas about how much effort they wanna put in. So sometimes it's hard to keep the peace and so.* [STD2]

Project Management with the Team. Students discussed getting organized, including the challenges of hybrid and remote work, coordinating schedules, the challenge in structuring work to ensure tasks were commenced early enough, and the various obstacles which interfered with their ability to plan effectively and set realistic goals for the team.

Supervision. In general, the reflections about supervision were positive, with comments about how they often discussed issues and milestones with them and how helpful supervisors were.

> *But I also have to say that the people were always there to give us feedback.* [STD4]

Though, some students disclosed how they sometimes found it difficult to disclose problems and that they were reluctant to use staff's time.

> *So a professor has much more important tasks than taking part in conversations between students who have no idea about a project.* [STD5]

Tool Use. Some of the students commented on the way they and their group used tools. Many related their tool use to those of other teams, highlighting how they can vary by size and culture of team. They also noted how groups sometimes struggled to use tools.

Use of Generative Artificial Intelligence. Many students reported that they did not use generative artificial intelligence (GAI), but they recognized the potential benefits of the technology. Often, the potential contributions to the design and planning stage were mentioned as were its ability to help level the playing field with aspects of interpersonal communication.

> *especially with robotics, one of the biggest things is merging a bunch of ideas that may already be pretty well advanced, so using generative AI can sometimes skip some steps. So I do see why people would want to do that.* [STD2]

> *For example, I see advantages in the fact that in both capstone projects I also worked with people who clearly had problems formulating things in [language] or in general.* [STD6]

They did, however, draw attention to many risks. These included obscuring authorship and other complications relating to intellectual property.

> *The disadvantage, on the other hand, is that it's not exactly clear which person did what, because each person always has their own style of writing.* [STD7]

They also included reflections on what might constitute cheating or otherwise give an appearance of improper contribution to a project. Some students were skeptical of the ability of GAI to accomplish difficult tasks.

> *Because we both worked very closely with research, for example, on some concepts or something that had just been published by the working groups or hadn't even been published yet and were therefore outside the scope of what any language models knew.* [STD6]

A few students mentioned that they were opposed to the use of the technology for creative tasks.

> *I don't really think there's much [of] a place in an academic setting when you're trying to learn how to do things yourself to get an AI to do things.* [STD1]

> *Even one step in that direction risks kind of losing what actually makes these things meaningful and beautiful.* [STD3]

Main Takeaways (Reflective Insights). Rather unsurprisingly given the theme of the interviews, a focus of the reflections was on transparency and perceived fairness of the assessments. Students also noted possible biases by instructors not having the full picture of the team dynamics and by the skewed visibility of individual contributions. Some students advocated the use of tools to be able to better document individual contributions, while the use of generative AI tools was considered to have both benefits and drawbacks.

7.2.2 Suggested Improvements. Students offered several suggestions on how to improve the capstone experience and address some of the challenges they experienced. Many of these focused on assessment and feedback, course structure, and supervision.

Assessment. Students called for assessment practices that were more transparent, scaffolded, and inclusive. They wanted clearer criteria and marking processes, with opportunities for staged feedback rather than reliance on high-stakes final submissions. A key concern was recognition of both visible outputs (e.g., code, reports) and less tangible contributions such as leadership or conflict management. Some suggested a "trust but verify" approach, where peer or self-reports are complemented by monitoring tools to ensure fairness without excessive oversight. Others recommended assessing collaboration explicitly as a learning outcome and giving students a limited choice in how individual and group marks are weighted. Overall, students emphasized fairness, recognition, and accountability in assessment.

> *[The assessor] like sat in and just silently watched us do our retrospective and I would personally think would prefer that to be a basis of marking.* [STD3]

> *So it would be nice if there was some kind of scheme on how to do it, to say, here, please make a protocol, maybe with guidelines, not concrete guidelines, but a pattern, so that it somehow looks like you all somehow record something and that it is the same for everyone and that everyone has to agree to it, [so] that it is openly visible to everyone and, yes, if it is officially specified, that you also have the guarantee that [...] your own performance is appreciated, but also the group performance and that everyone is [...] hopefully graded fairly or something.* [STD4]

Course Structure. Students suggested several improvements. Scaffolding of the course could be revised to guide the inexperienced and encourage, and facilitate, earlier writing-up. Project opportunities could more closely involve beneficiaries. Teams could be built on a stronger foundation, around ideas or selected by staff earlier in the course. Finally, scheduling could be improved by being

more flexible with team progress and adding more opportunities for contact with supervisors.

> *There should be some time aside, because most of the time the spec, the capstone project, and the deadlines just demand work. They demand something to be done, something to be made. It's kind of impossible to stop doing that, and it takes purposeful intent to put time aside to improve the team in a way.* [STD3]

Feedback. Suggestions from the students were focused on the foci, criticality, and platform for feedback, with a desire for more frequent and more individually focused feedback.

> *I think it would certainly be nice to have a kind of stable assessment at various points in the capstone project as to what grade the whole thing is heading towards.* [STD6]

Supervision. The students suggested that supervisors take a more active role in chasing disengaged peers, which included personally checking in with individual students to see how they were getting on with the project. They also advocated for closer supervision, including observations, to facilitate an open and consistent dialogue to better guide and support the team.

> *I think once every two weeks we had an hour-long session with the supervisor. It wasn't really enough and we weren't able to talk about what we really needed to have talked about. So I think having the supervisor be a more integral part of the project where they understand what's going on a bit more.* [STD1]

> *I believe like a simple one-on-one message just being like "Hey, how's it going? Could you comment more on the thing you're falling off or potentially, if someone's being too bossy or too bulldozerish?"* [STD2]

Main Takeaways (Suggested Improvements). Even if details about the assessment process were provided in the syllabus, students mentioned that even more details would be of help. Along the same lines, students explicitly asked for more scaffolding and explicit supervision. This seems to indicate that students ask for guidance both in the form of transparency regarding expectations and explicit supervision even though capstone projects are designed to provide more leeway than traditional courses.

7.2.3 Advice for Incoming Students. Students offered advice for other students based on their experience, to help set them up for success.

Helpful Mindset for Capstone Success. Among the advice were encouraging helpful mindsets. These included: being free, taking advantage of the freedoms and flexibility offered in capstone courses; being honest, to keep things clear with the team; being motivated, to make a commitment to undertake the project and to not take a capstone course only if you know you don't enjoy the work; and to prioritize the team over the grade.

> *But talk, talk, talk a lot with the supervisors, they are super nice, they help at every turn.* [STD4]

> *it's definitely important to put the team first to treat your own processes, to treat your own goals, and what you're doing as your own. Take ownership of that. And to in a way take on a mentality that you don't let anyone slow you down or stop you.* [STD3]

Recommended Work Practices. A number of recommendations were made, regarding working practices that future students might consider. These include: reading the brief; keeping minutes of meetings; setting out time to work on the project; communicating with staff and peers; holding each other accountable; and logging contributions.

> *I would also say document everything you do.* [STD4]

Strategic Planning and Tips. We considered that some of the advice constituted broader strategic planning and tips which students could consider. These included: reducing scope, to meet the needs of the capstone without putting unnecessary pressure on peers; and agreeing process and tools in advance, such that the team is clear on its expectations and workflows to facilitate a positive exchange between peers and to maximize productivity insofar as that is possible.

> *Don't develop a tunnel vision at the beginning of the capstone project for what you prepare yourself, in terms of subject matter and content, but can also follow the others, because otherwise you'll end up [...] with a horde of technical idiots and that's not really the point of the whole thing, you want to understand the project in general and you should make the commitment at the beginning to get a bit more of a holistic understanding and not just your own part.* [STD6]

> *You can't really influence how people assess you, but you can try to adapt a little bit to them if you know roughly how things will work, and see how you can work well according to the assessment scheme once it has been made transparent.* [STD8]

Main Takeaways (Advice for Incoming Students). Participants suggested that incoming students should develop an open-minded approach and use every opportunity to communicate with the instructors and external stakeholders. To support the workflow and the assessment process incoming students should get into the habit of journaling and documenting. Finally, being aware of their own contributions within the larger scope of the project and being open to adjust was mentioned as a successful strategy.

8 Stakeholder Perspectives

The extent to which a stakeholder becomes involved in assessment is one area we wanted to investigate in this study. In most cases, a university's assessment standard precludes external bodies from contributing to the summative assessment of a course. However, it is not unusual for a stakeholder to provide informal feedback to students, teams or instructors, and they can provide an essential professional, competency-based perspective on the design and methods of assessments as well as providing real-world context and motivation for students.

In this study, we aimed to survey and interview external stakeholders; however, we were unable to recruit a sufficient number to

conduct a meaningful online survey and were only able to carry out a small number of interviews. While this limits the depth of our insights, we believe that the consistency of responses provides a reasonably clear picture of how stakeholders are integrated into capstone projects. We were able to interview five stakeholders, as summarized in Table 12.

Table 12: Demographics of stakeholder interview participants

Participant ID	Gender	Continent	Years of Capstone Involvement
SKH1	Male	Europe	2 years
SKH2	Male	Europe	1 year
SKH3	Male	Europe	10 years
SKH4	Male	North America	10 years
SKH5	Male	Oceania	4 years

The thematic analyses of the five in-depth interviews with stakeholders still resulted in a number of themes: (1) Stakeholder motivation, (2) Extent and nature of involvement, (3) Input to assessment and feedback, (4) Engagement and alignment, (5) Challenges, and (6) Recommendations. More details regarding the different facets of statements assigned to these themes will be discussed in the following subsections.

8.1 Stakeholder motivation

The motivation for stakeholder involvement can be altruistic: interviewees spoke of helping to educate the next generation of software developers. Other common factors include recruitment - talent-spotting - and raising awareness of the company brand. Stakeholders propose projects which they either do not have the resource to pursue in-house, or that are speculative investigations that generate prototype ideas potentially leading to a more in-depth study.

8.2 Extent and nature of involvement

8.2.1 Stakeholder role. Stakeholders may take different roles, for example as a faculty member of one institution acting as a stakeholder for another institution. This can extend to acting as a "back-up" stakeholder in case the main company representative is not available.

Within that role, the following activities are common for stakeholders:

- gathering ideas and proposals from the company, checking their scope and suitability as projects
- offering up a range of projects for students to select from
- engaging in review of student 'pitches' to take a chosen project
- holding check-in meetings with students during the project, at a frequency varying from every 2 weeks to every 10 weeks, to give feedback on progress
- giving feedback to instructors, both during and at the close of the project
- attending closing showcase events and celebrating student achievement

As can be seen the majority of a stakeholder's role is secondary to assessment, but there is a common theme of contributing informally to formative and summative assessment.

8.3 Input to assessment and feedback

Stakeholders generally have no contribution to the formal grade, but have the opportunity to give feedback at various points during the project.

8.3.1 No formal contribution. Stakeholders observe that they are not asked for input, either get simply told what the marks are, or do not see the grades at all. In cases where they do have sight of a marking scheme, this is observed to be quite formal and scientific, but quite independent from them.

In cases where there is some input based, for example, on judging at a showcase event, the contribution of this input to the overall grade is also not spelled out. Stakeholders are generally happy with this arrangement, leaving the responsibility to the instructors who are hands-on with the students.

8.3.2 Informal feedback to instructors. Feedback to instructors may take place on a regular basis or at the end of the project, when a demonstration or showcase event is scheduled. The feedback at this stage is often just an informal chat with the instructors, and instructors expressed a preference for giving feedback more continuously. Stakeholders may not be aware whether or not the feedback contributes indirectly to the grade.

8.3.3 Formative feedback to teams. Stakeholders put more value on the feedback they give to teams, as this is seen to be valued by the teams themselves, particularly where given continuously to make suggestions and ensure teams are on the right lines. Some stakeholders like to give hints, for example about key decisions to make or how to provide more "wow factor", and see if teams have followed up on them at the next meeting. Feedback can be more formally part of the project schedule, at an initial design review or final showcase.

> *By this point, I think the project was actually done. So it wasn't a point of giving them feedback. It was just going, OK, let's just see what you've done.* [SKH2]

8.3.4 Trust relationship with teams. As some stakeholders have regular meetings with teams a different relationship emerges compared with the formal instructor relationship. Stakeholders note that they use techniques such as round-robin to make sure all team members can speak up, spend time informally chatting with teams. This can result in teams contacting the stakeholder independently for advice, or disclosing issues with teammates that they would not be comfortable reporting to the instructor.

8.3.5 Outcomes: Project showcase. Although they have little involvement with formal assessment, stakeholders are very engaged with the showcase or demonstration aspects of the capstone. Common themes here are the sense of excitement or buzz on showcase day, the opportunity to see what has been achieved. A good showcase presentation can make up for an underachieving project, and while some stakeholders feel that this element could be more formal,

others appreciate the disconnect from formal assessment which is offered by the award of prizes voted for by employers.

> *I'm biased, I want my teams to win!* [SKH1]

8.4 Engagement and alignment

8.4.1 Team dynamics. In their observations of teams, stakeholders identified some common behaviors including how vocal a student might be in meetings or presentations, compared with the amount of effort being put in. Structurally, a team assigned to backend development might be relatively introverted compared with the front-end team.

Stakeholders emphasized the importance of being a team player, and as part of that, being able to explain what part of the work they contributed to. They were also interested in who emerges as the leader, in order to inform instructors as well as to identify potential for graduate recruitment.

> *Real software has leads. Who's ultimately responsible for pulling everything together?* [SKH3]

8.4.2 Perspectives on real-world software development. Stakeholders are keen to communicate lessons and wisdom from their real-world experience. These include the importance of learning to deal with conflict, not expecting to work "alone in a cellar", and the ability to estimate what can be achieved in a given time frame. A consequence of effective estimation is being able to finish a task, and to understand the importance of having a build that can be demonstrated to an industry visitor at any point during the project.

These observations reflect the importance of *authenticity* in a project: representing how projects are managed in industry, and having real-world applicability.

> *It really is interesting how little they use the word estimate … in real business it's the first thing everybody says, well, how long is it going to take?* [SKH1]

> *Always have a build. You never know who's going to walk in the door.* [SKH3]

8.4.3 Employability. Stakeholder engagement with projects gives students more opportunities to develop employability skills and connect with employers, although a lot of students do not take advantage of these opportunities. Encouraging students to do the roles they are most comfortable with gives them a chance to showcase their skills. Encouraging students to use social media sites such as LinkedIn helps them to do this showcasing.

8.4.4 Differing priorities. Stakeholders observed that it is quickly apparent who the leader is, and who is just "hanging on". This has an impact on fairness of assessment, but in the main, external stakeholders are not concerned with this. Although (as noted above) it's important to know who did what, most important is the output of the team. It is to be expected in the real world that some team members will contribute more than others, and there can be different emphasis put on team versus individual contribution.

> *I care in business about what teams output. Individuals, less so.* [SKH1]

> *Life is the way group projects work … part of life is the output of the team is your output. And if you want to raise it, then you're going to have to go above and beyond to raise it* [SKH4]

8.4.5 Disagreements with instructors. It is unusual for stakeholders to challenge an assessment or final grade for a team or individual. Where this does occur, the instructor explains some of the science behind the grading and the stakeholder accepts this authority.

8.5 Challenges

8.5.1 Confidentiality and workload. The aim of providing an authentic project conflicts with IP and commercial sensitivities. This can mean that live projects or the "cool stuff" a team might hope to be able to investigate, cannot be made available or supported. Stakeholders also report challenges in allocating time to support students, and that there is not enough time in the project course schedule to give timely feedback.

8.5.2 Fairness of contribution. Student teams would ideally balance themselves to ensure a fair contribution, but a common observation is that two or three people are doing most of the work and that instructor support is needed to resolve this. If asked to comment on individual contribution, stakeholders find it very hard to judge; it can be misleading to look at GitHub commits, and if only one team is being observed by a stakeholder then they cannot compare expectations with other teams. Even the use of peer review can be taken as a routine activity by students rather than properly engaging. The weaker team members do not ask how they can contribute more to a team, in comparison with the stronger ones who are more concerned about this.

8.5.3 Use of Gen AI. While the opportunities for responsible use of Generative AI were not explicitly addressed, it was observed that it can give the weaker students less motivation to study. The availability of AI to students may make up for non-contributing students, leading to a more productive group. Availability of AI to stakeholders can reduce the need to engage with capstone courses, since a prototype project implementation can be generated in minutes.

> *I told him, we need a web interface for this and this. Maybe we can offer as a capstone project, and he came to me next day. I just did this one with AI, was really easy* [SKH5]

8.6 Recommendations

8.6.1 Reflections. Stakeholders reported that they're often amazed by the passion students have for the projects. They appreciate when they get positive feedback from the students.

8.6.2 Use of Gen AI. As with other aspects of the degree, capstones will need to change in order to deal with generative AI challenges, by putting the focus on other skills.

8.6.3 Communication. By letting students know that the stakeholder is in a supportive role rather than grading, teams can be more comfortable and happier to discuss a variety of issues.

8.6.4 Assessment. There are opportunities for more involvement in assessment: for example, providing evidence of engagement and contribution of students; viewing a mid-point presentation to

calibrate with other teams; seeing more of the final assessment, to compare with the brief that they set; having some influence on the final grade. While individual reports are helpful to assess contribution, they are not sufficient on their own.

8.6.5 Process. Stakeholders expressed a wish to get more involved in different parts of the process, and those who do have this involvement feel better informed. Where there are disagreements, some would like more visibility on the outcomes.

We found that few external stakeholders have any involvement with the formal assessment process. It is quite common for the involvement to be limited to setting or providing a number of projects for students to select, and then have no further involvement until a final showcase. Stakeholders expressed a wish to be more closely involved with other aspects.

9 Discussion

To discuss the findings from this working group, we return to the research questions and summarize the common themes, challenges, and solutions arising from the survey and interviews. In the conclusion section, we will compare our analysis with what was found in the literature review.

9.1 RQ1. Assessment methods and their effectiveness for evaluating contributions

Our quantitative data showed that both students and instructors reported a common core set of deliverables in capstone courses, with strong agreement on the importance of oral presentations, written reports, project demonstrations, and completed project builds. The frequencies are closely aligned, with over 80% of both groups identifying these as central components of assessment. Reflection essays and progress logs also showed similar patterns, reported by roughly 60% of instructors and students alike.

Where the groups differ is in the reporting of less common deliverables. Students rarely mentioned alternative forms of assessment beyond the standard set, with only 7% citing items such as peer feedback or peer reviews. In contrast, instructors more frequently pointed to additional assessment elements (27%), including peer evaluation, instructor observations, public demonstrations, code contributions, and video recordings. This suggests that while students primarily focus on the most visible, high-stakes deliverables that shape their grades, instructors recognize a broader range of evaluative practices—particularly those that capture process, performance, or professional engagement beyond the final product

The interviews provided more insight. A wide range of assessment methods was identified. Formative methods such as regular progress reviews and check-ins are effective in maintaining engagement and identifying individual contributions at a point where "free-riders" can be given a course-correction. Inclusion of git logs or tracking tools can also provide more data to aid fair assessment, as long as students accept that way of working. This requires students to understand that the correct use of collaborative version control is part of the expectations of the module. Summative assessments include graded demos, presentations, final reports, and the final product itself. Fair evaluation of contribution is typically achieved using a formula based on individual contribution (self-reported or instructor-based) and the mark for the final product, but this may be supplemented by metrics of process issues such as productivity and teamwork. Instructors spoke of using triangulation to come up with a final grade based on the various pieces of evidence, but also noted the amount of time this takes, and concerns about students "gaming the system" if they are aware of the formulae used, which can affect the effectiveness.

We asked both instructors and students about the extent to which different assessment types contributed to determining students' final grades. Across both groups in our quantitative data, instructor evaluation emerged as the dominant factor, with nearly all instructors and a majority of students rating it as highly significant. However, differences appear in the relative importance of other methods. Attitudes to peer evaluation in this data are divided, and this is reflected in interviews. Some instructors reported a positive impact on team dynamics from the open discussion of contributions, but others warned of possible fractured relationships and a tendency to award higher peer marks to the more popular and outgoing students. As a result, peer evaluation is often used as a formative support or to inform judgment rather than as part of the triangulation data for a summative assessment. While instructors placed very limited weight on self- and peer-assessment, students rated these somewhat higher, with about one third viewing peer evaluation as significant in our surveys, despite only 7% identifying peer evaluation as a capstone deliverable. External stakeholder feedback was consistently reported as the least influential by both groups, although students rated it slightly more highly than instructors. Taken together, these findings highlight broad agreement on the centrality of instructor evaluation, while also revealing that students perceive a greater significance for peer and stakeholder input than instructors report in practice.

9.2 RQ2. Instructor challenges for assessment

A major challenge for instructors is the workload taken by the range of assessment methods described above, particularly with giving timely feedback for formative exercises and handling the contribution-recording activities, as cohort sizes increase. The same issues also give rise to difficulties in consistency, higher numbers of conflicts, and more complex interactions with industry stakeholders. Assessing individual contributions is the other major challenge. These issues are reflected in the quantitative data and interviews.

Instructors agree that AI can be useful for students, but its use must be acknowledged. While AI can provide a useful first step in a literature review and accelerate student progress in this area, over-reliance can lead to shallow understanding, lack of skill development, originality or creativity, or inaccurate reports, for example, misrepresenting a student's contribution. In the extreme, where developed work is being used by an industry stakeholder, this might lead to legal issues.

9.3 RQ3. Student perceptions of fairness, transparency, and clarity

Student perceptions of assessment differ from those of instructors, for example, in how different aspects of the work, or input from external clients, contribute to the assessment. The difference may be explained by a reluctance of instructors to be explicit about these aspects to discourage gaming the system, as noted above.

Students are particularly concerned about how their contributions will affect their grade, and would like more transparency about assessment and contributions. This highlights a tension between instructor and student expectations. Transparent assessment is enhanced through the use of published rubrics, giving students the opportunity to assess their own performance [27] as well as understand how their work has been assessed. This has implications for assessment design. Keeping aspects such as the contribution formula confidential may be intended to encourage students to focus on the process rather than obsess about where to put most effort, but the consequential eroding of trust works against the wish for transparency and reduces the authenticity of the activity.

Other concerns about fairness can be related back to the issues of staff workload and resourcing. Students report that some assessors are not present, or only briefly present, at touchpoints during the project or the final assessment. This leads to concerns about an inconsistent picture across a cohort, with some assessors more engaged and with a better understanding than others. Extending this to feedback, students would prefer more "feed forward", i.e., opportunities for improvement in their next submission, and observed a lack of critical feedback in peer reviews. This reflects a concern about peer evaluation and team dynamics, with friendships sometimes leading to biased assessments.

Students also raised GenAI as a potential challenge to fairness and transparency, reflecting the concerns of instructors. They were, however, largely neutral about the impact of GenAI on fairness and transparency.

9.4 RQ4. External stakeholders and their role in assessment

Stakeholder contribution to assessment is generally more limited than as perceived by students, with many institutions prohibiting non-faculty members from contributing to the summative assessment. Stakeholder perception of fairness is based on their observations of how team members conduct themselves, but they generally are not concerned with the triangulation that leads to a mark. Stakeholder involvement is motivated differently from instructors, as they are interested in recruitment, brand awareness, and exploring technical challenges that they don't have the resources to investigate in-house.

AI is recognized as a game-changer - perhaps in surprising ways, as the specification of small technical challenges can as easily be explored internally with the help of AI rather than being set as a creative team challenge.

9.5 RQ5. Evidence-based strategies to address challenges

Instructors identified integrating self- and peer-assessment components, with robust calibration, as a common strategy, along with regular formative assessments and feedback, use of contribution-tracking tools, and better advice and training on AI. However, each of these strategies risks increasing workload, which is identified as one of the most critical factors in implementing new strategies. The use of tools and AI guidance may be more of a front-load cost, which will lead to reduced workload further down the line, if used strategically.

The development of rubrics that are detailed and flexible was also raised by instructors as a strategy to address challenges. Student buy-in is a critical factor, but this does conflict to some extent with instructor preference for keeping more detailed aspects of the assessment methodology undisclosed (such as contribution calculation formulae), to avoid students "gaming the system" or working to the assessment criteria rather than simply engaging as best they can with the capstone project. Appropriate choices need to be made in design of assessments and rubrics in order to provide sufficient transparency for students.

Student suggestions to address challenges are focused on having more guidance, scaffolding, more frequent feedback, and more active supervision - challenging the workload and resource issue further. Stakeholders also suggest improving communication so that students know what the role of the stakeholder is, and so they understand that an honest interaction will not affect their assessment.

10 Threats to Validity

We encountered some limitations in designing, executing, and analyzing the results of our study. To begin with, we observed differences in terminology due to the range of contexts and cultures across all the contributors to the study. This could lead to differing interpretations, which we mitigated by establishing common terminology as discussed in Section 2.

Because the survey was administered anonymously, it was not possible to verify whether individual participants completed it multiple times, representing a potential threat to internal validity.

Furthermore, selection bias may have been introduced through the inclusion of interview participants known to the researchers, as these individuals may differ systematically from those without prior connections to the research team. With three different sets of interviews taking place (students, instructors, stakeholders), we did not use a standard coding taxonomy. Anonymised transcripts were independently coded by WG members, and codes were then merged to form a common code book for each of the three sets. Subgroups of the WG reviewed the transcripts in each set to reduce the risk of missing codes.

While there was plenty of participation in interviews of students and instructors, the number of stakeholders was low ($N = 5$) and had a limited geographical spread. No stakeholders completed the quantitative survey. The impact of this may be that our conclusions about the limited role of stakeholders in assessment could be challenged. However, it is consistent with observations from interviews and surveys of instructors and students, who also noted limited input to assessment. There is also some potential for bias, as at least two of the interviewed stakeholders also had experience as instructors.

Another internal threat comes from the fact that we cannot assess the amount of overlap between interview and survey participants. We did encourage all interviewees to also complete the corresponding survey, but are not privy to whether or not they did so. While having more information about this would strengthen the analysis of how much interview and survey data align, such additional keeping track of participants would have threatened the integrity of the individual data submitted: students might not have

consented to fill out the survey if they knew that their data would be correlated with interview data.

The hierarchy between students and instructors was the source for another potential threat to internal validity. To mitigate this, in one case the researcher who was an instructor as well offered candidates to be interviewed by a third party who had been trained w.r.t. the interview guidelines. This person then transcribed and anonymized the interviews before giving the researcher access to the interview data.

Finally, the geographic spread of the interviewers and interviewees is both a strength and a weakness of the participant selection process. On one hand, it strengthens the findings because of the institutional and cultural diversity. On the other hand, it weakens the findings because there might be an unintended bias to a specific institution's curricular decisions, including, but not limited to, class sizes and project types, which then happen to be "representative" for the geographic region.

As a result, there is a limitation to the generalizability of the results, which should be borne in mind when interpreting the findings. That said, we firmly believe that the breadth of aspects discovered through our analyses offers starting points for more in-depth studies conducted within any more narrowly defined set of participants considered in future work.

11 Conclusion

Our working group set out to understand how team-based capstone projects are currently assessed in computing programs, what challenges different stakeholders encounter, and which strategies show promise for improving transparency, fairness, and workload from both instructors' and students' perspectives. Drawing on a multi-institutional survey (N=66) and thirty in-depth interviews with students (N=12), instructors (N=13), and external stakeholders (N=5), our findings offer a coherent picture that addresses each of our five research questions — current assessment practices (RQ1), instructor challenges (RQ2), student perceptions (RQ3), stakeholder roles (RQ4), and evidence-based recommendations (RQ5).

Our findings from the surveys and interviews are generally consistent with those in the literature review, in terms of instructor perspective. It's evident that the adoption of practices for assessment remains varied and inconsistent, and instructors are reluctant to lean heavily on techniques for individual or team contributions to provide a major part of the triangulated assessment, regarding them as not reliable or scalable.

We identified a gap in the literature when it came to the student and stakeholder perspective. We have addressed this from the student perspective, less successfully for the stakeholders. As noted in the discussion, perceptions of fairness are seen in a desire by students for more transparency in the formulae and contribution measures, and generally more feedback and guidance. Student ambivalence about peer evaluation and potential for bias was identified strongly, with a desire for more accurate and objective capture of their individual efforts.

The practical challenge of workload has to be addressed in considering how to implement improved assessment strategies. The balance of workload with fair assessment was identified as a key question from the literature review. Evidence from our survey and interviews shows that approaches to tracking contributions, combined with the desire from students to have more ongoing formative feedback and guidance, tip the balance toward a higher workload. For improvements like these to be feasible, the workload barrier needs to be overcome. Automation and AI may be part of a solution, if concerns over perceptions of fairness can be convincingly overcome.

One area that stood out in our review of the literature was the lack of discussion of GenAI challenges. This may be a consequence of the timescale of publication and the accelerated pace of developments in this area. Challenges and some emerging practices were identified in the surveys, but a clear picture is still to develop: many instructors are ambivalent about its use, but both students and instructors see a need for clear guidance on what is, and is not, acceptable. The need for unambiguous institutional policies on AI use in teaching and assessment is well-established [37], and the picture is improving although policies may focus on textual written work. The nature of computing assessments, covering written work, code and other artifacts, requires computing departments to take a lead in co-creating policies alongside stakeholders and students, for use of GenAI across all assessments - with application to the specific challenges of capstone projects.

Capstone projects will remain a signature culminating experience for computing graduates. By triangulating the perspectives of students, instructors, and external stakeholders, our work provides an insight into factors that affect the fairness and rigour of assessment methods, thereby better preparing students for the collaborative realities of the professional software industry.

References

[1] Robert Adams and Carsten Kleiner. 2016. Collaboration support in an international computer science capstone course. In *Social Computing and Social Media: 8th International Conference, SCSM 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17–22, 2016. Proceedings 8*. Springer, 313–323.

[2] Ehsan Ahmad and Ali Alammary. 2022. Towards unified management of software capstone projects in Saudi universities: a survey-based study. *Arab Gulf Journal of Scientific Research* 40, 2 (2022), 118–138. https://doi.org/10.1108/AGJSR-04-2022-0037

[3] Ray Bareiss and Edward Katz. 2011. An exploration of knowledge and skills transfer from a formal software engineering curriculum to a capstone practicum project. In *2011 24th IEEE-CS Conference on Software Engineering Education and Training (CSEE&T)*. IEEE, IEEE Computer Society, Los Alamitos, CA, 71–80. https://doi.org/10.1109/CSEET.2011.5876159

[4] Adelina Basholli, Fesal Baxhaku, Dimitris Dranidis, and Thanos Hatziapostolous. 2013. Fair assessment in software Engineering Capstone Projects. In *Proceedings of the 6th Balkan Conference in Informatics (BCI '13)*. ACM, New York, NY, 244–250. https://doi.org/10.1145/2490257.2490268

[5] María Cecilia Bastarrica, Daniel Perovich, and Maíra Marques Samary. 2017. What Can Students Get from a Software Engineering Capstone Course?. In *2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET)*. IEEE, IEEE Computer Society, Los Alamitos, CA, 137–145. https://doi.org/10.1109/ICSE-SEET.2017.15

[6] Steven Bradley, Miranda C Parker, Rukiye Altin, Lecia Barker, Sara Hooshangi, Thom Kunkeler, Ruth G Lennon, Fiona McNeill, Julià Minguillón, Jack Parkinson, et al. 2023. Modeling Women's Elective Choices in Computing. In *Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education*. ACM, New York, NY, 196–226.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Quincy Brown, Frank Lee, and Suzanne Alejandre. 2009. Emphasizing soft skills and team development in an educational digital game design course. In *Proceedings of the 4th international Conference on Foundations of Digital Games*. ACM, New York, NY, 240–247. https://doi.org/10.1145/1536513.1536557

[9] Håkan Burden, Jan-Philipp Steghöfer, and Oskar Hagvall Svensson. 2019. Facilitating Entrepreneurial Experiences through a Software Engineering Project Course. In *2019 IEEE/ACM 41st International Conference on Software Engineering:*

Software Engineering Education and Training (ICSE-SEET). IEEE Computer Society, Los Alamitos, CA, 28–37. https://doi.org/10.1109/ICSE-SEET.2019.00012

[10] Joao Alvaro Carvalho, Rui Dinis Sousa, and Jorge Oliveira Sá. 2010. Information Systems Development Course: Integrating Business, IT and IS Competencies. In *2010 IEEE Transforming Engineering Education: Creating Interdisciplinary Skills for Complex Global Environments*. IEEE Computer Society, Los Alamitos, CA, 1–20. https://doi.org/10.1109/TEE.2010.5508847

[11] María Cecilia Bastarrica, Francisco J. Gutierrez, María Marques, and Daniel Perovich. 2023. On the Impact of Grading on Teamwork Quality in a Software Engineering Capstone Course. *IEEE Access* 11 (2023), 36492–36503. https://doi.org/10.1109/ACCESS.2023.3265302

[12] Showkat Chowdhury, Abdul Jalloh, Ruben Rojas-Oviedo, Mohamed Seif, and Amir Mobasher. 2007. Capstone Design Course as a Tool for Assessment and Improvement. In *2007 Annual Conference & Exposition*. 12–342.

[13] John W Creswell and J David Creswell. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publishing, Thousand Oaks, CA.

[14] Denny Davis, Steven Beyerlein, Olakunle Harrison, Phillip Thompson, and Michael Trevisan. 2007. Assessments for three performance areas in capstone engineering design. In *2007 Annual Conference & Exposition*. 12–293.

[15] Denny Davis and Peter Rogers. 2015. Structuring capstone design assessment to achieve student, faculty, and employer priorities. In *2015 ASEE Annual Conference & Exposition*. 26–1409.

[16] Morton Deutsch. 1949. A theory of co-operation and competition. *Human relations* 2, 2 (1949), 129–152.

[17] Nathan Eloe and Charles Hoot. 2020. Accommodating Shortened Term Lengths in a Capstone Course using Minimally Viable Prototypes. In *2020 IEEE Frontiers in Education Conference (FIE)*. IEEE, IEEE Computer Society, Los Alamitos, CA, 1–8. https://doi.org/10.1109/FIE44824.2020.9273821

[18] Vivienne Farrell, Gilbert Ravalli, Graham Farrell, Paul Kindler, and David Hall. 2012. Capstone project: fair, just and accountable assessment. In *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 168–173. https://doi.org/10.1145/2325296.2325339

[19] Martin R Fellenz. 2006. Toward fairness in assessing student groupwork: A protocol for peer evaluation of individual contributions. *Journal of Management Education* 30, 4 (2006), 570–591.

[20] Ridi Ferdiana. 2020. The Triangulation Assessment Model for Capstone Project in Software Engineering. In *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE Computer Society, Los Alamitos, CA, 131–134. https://doi.org/10.1109/ICITEE49829.2020.9271781

[21] Henrik Gustavsson and Marcus Brohede. 2019. Continuous assessment in software engineering project course using publicly available data from GitHub. In *Proceedings of the 15th International Symposium on Open Collaboration*. ACM, New York, NY, 1–6. https://doi.org/10.1145/3306446.3340820

[22] Nicole Herbert. 2018. Reflections on 17 years of ICT Capstone Project Coordination: Effective Strategies for Managing Clients, Teams and Assessment. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) *(SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 215–220. https://doi.org/10.1145/3159450.3159584

[23] Sara Hooshangi, Asma Shakil, Subhasish Dasgupta, Karen C. C. Davis, Mohammed Farghally, KellyAnn Fitzpatrick, Mirela Gutica, Ryan Hardt, Steve Riddle, and Mohammed Seyam. 2025. Instructors' Perspectives on Capstone Courses in Computing Fields: A Mixed-Methods Study. In *2024 Working Group Reports on Innovation and Technology in Computer Science Education* (Milan, Italy) *(ITiCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 68–94. https://doi.org/10.1145/3689187.3709608

[24] Bowen Hui. 2024. Guiding Principles for Assessing Software Engineering Teams. In *2024 IEEE Frontiers in Education Conference (FIE)*. IEEE Computer Society, Los Alamitos, CA, 1–9. https://doi.org/10.1109/FIE61694.2024.10893426

[25] Cruz Izu. 2018. Improving Outcomes for a Masters Capstone IT Project. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. IEEE Computer Society, Los Alamitos, CA, 301–307. https://doi.org/10.1109/TALE.2018.8615268

[26] David W Johnson and Roger T Johnson. 2005. New developments in social interdependence theory. *Genetic, social, and general psychology monographs* 131, 4 (2005), 285–358.

[27] Anders Jonsson. 2014. Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education* 39, 7 (2014), 840–852. https://doi.org/10.1080/02602938.2013.875117

[28] An Ju, Adnan Hemani, Yannis Dimitriadis, and Armando Fox. 2020. What Agile Processes Should We Use in Software Engineering Course Projects?. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, 643–649. https://doi.org/10.1145/3328778.3366864

[29] Jayden Khakurel and Jari Porras. 2020. The Effect of Real-World Capstone Project in an Acquisition of Soft Skills among Software Engineering Students. In *2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE&T)*. IEEE, IEEE Computer Society, Los Alamitos, CA, 1–9. https://doi.org/10.1109/CSEET49119.2020.9206201

[30] Phillip A Laplante, Joanna F Defranco, and Everton Guimaraes. 2019. Evolution of a graduate software engineering capstone course—a course review. *Int. J. Eng. Educ* 35, 4 (2019), 999–1007.

[31] Timoth C Lederman. 2010. Evolution of capstone-courses in software engineering a finishing school. *Journal of Computing Sciences in Colleges* 25, 6 (2010), 129–135.

[32] Deepan Lobo, Drashti Patel, Jorim Morainvile, Prateek Shekhar, and Pramod Abichandani. 2021. Preparing Students for Drone Careers Using Active Learning Instruction. *IEEE Access* 9 (2021), 126216–126230. https://doi.org/10.1109/ACCESS.2021.3110578

[33] Alejandra J. Magana, Ying Ying Seah, and Paul Thomas. 2018. Fostering Cooperative Learning with Scrum in a Semi-Capstone Systems Analysis and Design Course. *Journal of Information Systems Education* 29, 2 (2018), 75–92. https://aisel.aisnet.org/jise/vol29/iss2/4

[34] Maira Marques, Sergio F Ochoa, Maria Cecilia Bastarrica, and Francisco J Gutierrez. 2017. Enhancing the student learning experience in software engineering project courses. *IEEE Transactions on Education* 61, 1 (2017), 63–73.

[35] Larry J McKenzie, Michael S Trevisan, Denny C Davis, and Steven W Beyerlein. 2004. Capstone design courses and assessment: A national study. In *Proceedings of the 2004 American Society of Engineering Education Annual Conference & Exposition*. 1–14.

[36] Alexander Mitchell, Michael James Scott, Bedour Alshaigy, Wendell Bento Geraldes, Rita Garcia, Mirela Gutica, Hieke Keuning, Ellie Lovellette, PD Parthasarathy, Seán Russell, et al. 2024. Exploring Approaches to Assessing Student Teamwork in Undergraduate Computing Projects. In *Proceedings of the 2024 on ACM Virtual Global Computing Education Conference V. 2*. ACM, New York, NY, 285–286.

[37] Benjamin Luke Moorhouse, Marie Alina Yeo, and Yuwei Wan. 2023. Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open* 5 (2023), 100151. https://doi.org/10.1016/j.caeo.2023.100151

[38] Mohamed M. Morsy, Md. N. Uddin, and Faycal Znidi. 2024. Multidisciplinary Capstone Senior Design Projects: A Comparative Analysis of Industry–Sponsored and Faculty–Sponsored Projects Using Comprehensive Performance Metrics. *Eductation Sciences* 14, 3 (2024), 314. https://doi.org/10.3390/educsci14030314

[39] Bahram Nassersharif and Carl-Ernst Rousseau. 2010. Best practices in assessing capstone design projects. In *proceedings of the 2010 capstone design conference*.

[40] Andres Neyem, Jose I. Benedetto, and Andres F. Chacon. 2014. Improving software engineering education through an empirical approach: lessons learned from capstone teaching experiences. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (Atlanta, Georgia, USA) *(SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 391–396. https://doi.org/10.1145/2538862.2538920

[41] Andres Neyem, Jose Carrasco-Aravena, Alison Fernandez-Blanco, and Juan Pablo Sandoval Alcocer. 2025. Exploring the Adaptability and Usefulness of Git-Truck for Assessing Software Capstone Project Development. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1* (Pittsburgh, PA, USA) *(SIGCSETS 2025)*. Association for Computing Machinery, New York, NY, USA, 847–853. https://doi.org/10.1145/3641554.3701798

[42] Duc Man Nguyen, Tien Vu Truong, and Nguyen Bao Le. 2013. Deployment of capstone projects in software engineering education at duy tan university as part of a university-wide project-based learning effort. In *2013 Learning and Teaching in Computing and Engineering*. IEEE, IEEE Computer Society, Los Alamitos, CA, 184–191.

[43] Maria Paasivaara, Casper Lassenius, Daniela Damian, Petteri Räty, and Adrian Schröter. 2013. Teaching students global software engineering skills using distributed Scrum. In *2013 35th International Conference on Software Engineering (ICSE)*. IEEE, IEEE Computer Society, Los Alamitos, CA, 1128–1137. https://doi.org/10.1109/ICSE.2013.6606664

[44] Maria Paasivaara, Jari Vanhanen, and Casper Lassenius. 2019. Collaborating with industrial customers in a capstone project course: the customers' perspective. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, IEEE Computer Society, Los Alamitos, CA, 12–22.

[45] Attiqa Rehman and Monica Divitini. 2023. Categorisation of Software Engineering Capstone Projects. *NIKT: Norsk IKT-konferanse for forskning og utdanning* 4 (2023), 14 pages. https://www.ntnu.no/ojs/index.php/nikt/article/view/5789 Vitenskapelig artikkel.

[46] José Carlos Sánchez-Prieto, Adriana Gamazo, Juan Cruz-Benito, Roberto Therón, and Francisco J. García-Peñalvo. 2020. AI-Driven Assessment of Students: Current Uses and Research Trends. In *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences*, Panayiotis Zaphiris and Andri Ioannou (Eds.). Springer International Publishing, Cham, 292–302.

[47] Sangit Sasidhar and Jithin Vachery. 2024. Leveraging Industry 4.0 in Education for Remote Implementation in a Team-Based Computer Engineering Capstone Project. In *2024 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.

[48] Ruth Schorr. 2020. Experience Report on Key Success Factors for Promoting Students' Engagement in Software Development Group Projects. In *2020 IEEE*

World Conference on Engineering Education (EDUNINE). IEEE Computer Society, Los Alamitos, CA, 1–5. https://doi.org/10.1109/EDUNINE48860.2020.9149536

[49] Asma Shakil and Paul Denny. 2024. Enhancing Student Engagement in Large-Scale Capstone Courses: An Experience Report. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*. ACM, New York, NY, 722–728. https://doi.org/10.1145/3649217.3653580

[50] Li Ze Shi, Arony Nowshin Nawar, Devathasan Kezia, and Damian Daniela. 2023. Software is the Easy Part of Software Engineering - Lessons and Experiences from A Large-Scale, Multi-Team Capstone Course. In *Proceedings of the 45th International Conference on Software Engineering: Software Engineering Education and Training* (Melbourne, Australia) *(ICSE-SEET '23)*. IEEE Computer Society, Los Alamitos, CA, 223–234. https://doi.org/10.1109/ICSE-SEET58685.2023.00027

[51] Maria Spichkova. 2019. Industry-Oriented Project-Based Learning of Software Engineering. In *2019 24th International Conference on Engineering of Complex Computer Systems (ICECCS)*. IEEE Computer Society, Los Alamitos, CA, 51–60. https://doi.org/10.1109/ICECCS.2019.00013

[52] Jan-Philipp Steghöfer, Håkan Burden, Regina Hebig, Gul Calikli, Robert Feldt, Imed Hammouda, Jennifer Horkoff, Eric Knauss, and Grischa Liebel. 2018. Involving External Stakeholders in Project Courses. *ACM Trans. Comput. Educ.* 18, 2, Article 8 (July 2018), 32 pages. https://doi.org/10.1145/3152098

[53] Mark Steiner, Junichi Kanai, Cheng Hsu, Richard Alben, and Lester Gerhardt. 2011. Holistic assessment of student performance in multidisciplinary engineering capstone design projects. *International Journal of Engineering Education* 27, 6 (2011), 1259.

[54] Ben Stephenson, Mark James, Nigel Brooke, and John Aycock. 2016. An Industrial Partnership Game Development Capstone Course. In *Proceedings of the 17th Annual Conference on Information Technology Education*. ACM, New York, NY, 136–141. https://doi.org/10.1145/2978192.2978214

[55] Anya Tafliovich, Francisco Estrada, and Thomas Caswell. 2019. Teaching software engineering with free open source software development: An experience report. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. ScholarSpace, Honolulu, HI, 7731–7741. https://doi.org/10.24251/HICSS.2019.931

[56] Tuukka M Takala, Lauri Malmi, Roberto Pugliese, and Tapio Takala. 2016. Empowering students to create better virtual reality applications: A longitudinal study of a VR capstone course. *Informatics in Education* 15, 2 (2016), 287–317.

[57] Joo Tan and Mark Jones. 2008. A case study of classroom experience with client-based team projects. *Journal of Computing Sciences in Colleges* 23, 5 (2008), 150–159.

[58] Charles C Tappert and Allen Stix. 2011. A Decade Review of a Masters-Level Real-World-Projects Capstone Course. In *Proc. Info. Systems Educators Conf. (ISECON 2011)*, Vol. 28. Education Special Interest Group of the AITP, 13 pages.

[59] Saara Tenhunen, Tomi Männistö, Matti Luukkainen, and Petri Ihantola. 2023. A Systematic Literature Review of Capstone Courses in Software Engineering. *Information and Software Technology* 159 (2023), 107191. https://doi.org/10.1016/j.infsof.2023.107191

[60] Michael Trevisan, Denny Davis, Steven Beyerlein, Phillip Thompson, and Olakunle Harrison. 2006. A review of literature on assessment practices in capstone engineering design courses: Implications for formative assessment. In *2006 Annual Conference & Exposition*. 11–112.

[61] Laura Tubino, Andrew Cain, Jean-Guy Schneider, Dhananjay Thiruvady, and Niroshinie Fernando. 2020. Authentic individual assessment for team-based software engineering projects. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering Education and Training* (Seoul, South Korea) *(ICSE-SEET '20)*. Association for Computing Machinery, New York, NY, USA, 71–81. https://doi.org/10.1145/3377814.3381702

[62] Laura Tubino, Jean-Guy Schneider, Andrew Cain, Dhananjay Thiruvady, and Chathu Ranaweera. 2021. Reforming Assessment: Challenges Beyond Design. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE Computer Society, Los Alamitos, CA, 78–88. https://doi.org/10.1109/ICSE-SEET52601.2021.00017

[63] Maria Vasilevskaya, David Broman, and Kristian Sandahl. 2015. Assessing large-project courses: model, activities, and lessons learned. *ACM Transactions on Computing Education (TOCE)* 15, 4 (2015), 1–30.

[64] Michael Verdicchio. 2021. Hurricanes and Pandemics: An Experience Report on Adapting Software Engineering Courses to Ensure Continuity of Instruction. *Journal of Computing Sciences in Colleges* 36, 5 (jan 2021), 150–159.

[65] Brian R von Konsky and Jim Ivins. 2008. Assessing the capability and maturity of capstone software engineering projects. In *Proceedings of the tenth conference on Australasian computing education*, Vol. 78. Australian Computer Society, Sydney, 171–180.

[66] Yani Widyani. 2013. Experience in Software Development Project Course. *Procedia Technology* 11 (2013), 1018–1026.

[67] Rafal Włodarski, Aneta Poniszewska-Marańda, and Jean-Remy Falleri. 2022. Impact of software development processes on the outcomes of student computing projects: A tale of two universities. *Information and Software Technology* 144 (2022), 106787. https://doi.org/10.1016/j.infsof.2021.106787

[68] Jingxiao Zhang, Zhiyue Zhang, Simon P. Philbin, Henk Huijser, Qian Wang, and Ruoyu Jin. 2022. Toward next-generation engineering education: A case study of an engineering capstone project based on BIM technology in MEP systems. *Computer Applications in Engineering Education* 30, 1 (jan 2022), 146–162. https://doi.org/10.1002/cae.22448

[69] Stephen J. Zilora. 2015. Industry-Emulated Projects in the Classroom. In *Proceedings of the 16th Annual Conference on Information Technology Education (SIGITE '15)*. ACM, New York, NY, 115–119. https://doi.org/10.1145/2808006.2808029

Appendix

Table 13: List of survey questions for Instructors

ID	Question text	Answer format
Q1	If you would like to participate in this survey and you are 18 years of age, click yes to begin or no to exit.	Multiple Choice: Yes; No
	Section 1: Course Background Information	
Q2	Please choose the capstone course level that you have been involved with (choose a level; if you have more than one, you can fill this form again).	Multiple Choice: Two year program (associate or diploma); Three or four year program (bachelor's); Graduate program (master's)
Q3	What is the duration of your capstone course?	Multiple Choice: One Semester; Two Semesters; Other
Q4	What is the typical class size for your capstone courses?	Multiple Choice: Less than 20; 20–40; 40–100; 100–200; More than 200
Q5	Are your capstone projects typically:	Multiple Choice: Team-based; Individual; A mix of both; Other
Q6	What is the typical size of students' teams in your capstone course?	Multiple Choice: 1–2; 2–3; 3–4; 4–5; 5–6; Greater than 6
	Section 2: Existing Assessment Practices	
Q7	Which of the following assessment approaches do you use when evaluating capstone projects?	Multiple Choice: Formative assessment only (ongoing evaluation throughout the course); Summative assessment only (end); A combination of both
Q8	Which of the following best describes your approach to assessing capstone projects?	Multiple Choice: Primarily subjective (e.g., holistic judgment, open-ended feedback); Primarily objective (e.g., rubric-based scoring, standardized criteria); A combination of both; Other
Q9	What grading scheme do you use in evaluating individual grades for team members?	Multiple Choice: Same grade for all team members; Individual grades based on instructor assessment; Individual grades based solely on team evaluation; Combination of instructor/stakeholder and team evaluation; Other
Q10	Which deliverables counted towards the final grade in your capstone course? (Select all that apply)	Multiple Choice (Multi-Select): Written reports; Oral presentations; Reflection essays or journals; Project demonstrations; Completed project build/deliverable; Progress meeting minutes or team logs; Other
Q11	To what extent does each of the following assessment types contribute to determining students' final grades in your capstone course? (Self-assessment; Peer evaluation; Instructor evaluation; External stakeholder feedback)	Multiple Choice: 5-item Likert-type scale
Q12	Which of the following do you rely on when grading or assessing capstone projects? (Select all that apply)	Multiple Choice (Multi-Select): TAs; Other instructors/co-lecturers; External stakeholders (e.g., industry experts, clients); Graduate students (not designated as TAs); I handle all grading independently; Other
Q13	How would you rate the adequacy of the grading/marking support you receive to help you assess the capstone project?	Multiple Choice: 5-item Likert-type scale
Q14	In your experience, how effective are the following methods in evaluating team contribution? (Peer evaluation; Self-assessment; Instructor evaluation; External stakeholder evaluation)	Multiple Choice (Grid): 6-item Likert-type scale
Q15	In your experience, how effective are the following methods in evaluating individual contribution? (Peer evaluation; Self-assessment; Instructor evaluation; Written reports/documentation; Oral presentations/demos; External stakeholder evaluation)	Multiple Choice (Grid): 6-item Likert-type scale
	Section 3: Assessment Challenges	
Q16	Please elaborate on any limitations you have experienced in using peer evaluations for assessing individual contribution.	Text Answer

Continued on next page

Table 13 (continued)

ID	Question text	Answer format
Q17	Please elaborate on any limitations you have experienced in using external stakeholder evaluation for assessing individual and/or team contribution.	Text Answer
Q18	What are the key challenges you encounter when assessing team-based projects? (Select top 3)	Multiple Choice (Multi-Select): Difficulty distinguishing individual contributions; Bias in peer evaluations; Time constraints; Conflict in team; Lack of grading/marking support; Impact of generative AI tools; Other
Section 4: Use of Generative AI		
Q19	Have you encountered the use of generative AI tools (e.g., ChatGPT) during the capstone project process?	Multiple Choice: Yes, extensively; Yes, occasionally; No, not at all; Unsure
Q20	On a scale of 1 (No Impact) to 5 (Significant Impact), to what extent do you think generative AI tools affect the overall quality of capstone projects?	Multiple Choice: 5-item Likert-type scale
Q21	Are there clear guidelines or policies at your institution regarding the appropriate use of generative AI for academic work?	Multiple Choice: Yes, clearly stated; Somewhat unclear; No, not at all; Not sure
Q22	Rate your agreement with the statement: "The use of Generative AI tools by students compromises the fairness and integrity of capstone assessments."	Multiple Choice: 5-item Likert-type scale
Q23	Have you adjusted your assessment methods or rubrics/guidelines to account for possible use of generative AI by students?	Multiple Choice: Yes, significantly; Yes, slightly; No, but I plan to; No, and I do not plan to; Other
Q24	Do you use any strategies or tools to detect AI-generated content in student submissions?	Multiple Choice: Yes, specialized software; Yes, manual checks (e.g., style inconsistencies); No, I do not currently check; Other
Q25	In your opinion, what is the long-term effect of GenAI tools on the design and assessment of your capstone course? Both positive and negative effects can be discussed.	Text Answer
Section 5: Proposing Solutions		
Q26	Which of the following strategies could most effectively address the challenges you face in assessing team-based projects? (Select all that apply)	Multiple Choice (Multi-Select): Detailed/dynamic rubrics; Regular formative/iterative feedback; Calibrated self-/peer-assessment; Digital tools to track individual contributions; Training on AI ethics/usage; Additional resources; Other
Q27	What are the most critical factors to consider when implementing new assessment strategies in your courses? (Select your top three)	Multiple Choice (Multi-Select): Scalability; Transparency; Time and resource availability; Student acceptance and training; Alignment with learning outcomes; Other
Q28	What measures or assessment designs could better balance the evaluation of individual contributions and overall team performance?	Text Answer
Q29	Based on your experience, what modifications or improvements would you suggest to enhance the fairness and transparency of capstone project assessments?	Text Answer
Q30	How can emerging technologies (e.g., generative AI, analytics tools) be utilized to improve the assessment process in capstone projects?	Text Answer
Q31	What innovative approaches or practices would you recommend for future capstone assessments to ensure both rigorous evaluation and student development?	Text Answer
Section 6: Demographics		
Q32	Your department or academic unit (CS, Informatics,etc.)	Text Answer
Q33	Country	Text Answer
Q34	Role in the Capstone Course	Multiple Choice (Multi-Select): Course Coordinator; Instructor/Teacher; Teaching Assistant; Other

Continued on next page

Table 13 (continued)

ID	Question text	Answer format
Q35	How many years have you been teaching capstone courses?	Multiple Choice: Less than 2 years; 2–5 years; 6–10 years; More than 10 years; Other

Table 14: List of survey questions for Students

ID	Question text	Answer format
Q1	If you would like to participate in this survey and you are 18 years of age, click yes to begin or no to exit.	Multiple Choice: Yes; No
	Section 1: Course Background Information	
	Please answer the questions in this section with **one specific capstone course** in mind- preferably the most recent team-based capstone you completed.	
Q2	Was your capstone project completed as:	Multiple Choice: A team project; An individual project; Other
Q3	At what level was your capstone course?	Multiple Choice: Two year program (associate or diploma); Three or four year program (bachelor's); Graduate program (master's)
Q4	Including yourself, how many students were on your capstone project team? (This refers specifically to the group you worked with on the capstone project—not the total number of students in the course.)	Multiple Choice: 2; 3; 4; 5; 6; Greater than 6
Q5	How was your team formed? (Select all that apply) This refers to how members were grouped into teams for the capstone project.	Multiple Choice (Multi-Select): Self-chosen; Instructor-Assigned; Random; Other
	Section 2: Assessment Elements in Your Capstone Course	
Q6	Which deliverables counted toward your final grade? (Select all that apply)	Multiple Choice (Multi-Select): Written reports (e.g., final documentation, technical report); Oral presentations (e.g., progress presentation, final pitch, final presentation); Reflection essays or journals; Project demonstrations (e.g., live demo, walkthrough); Completed project build or deliverable; Progress meeting minutes or team logs; Other
Q7	For each evaluation method, indicate how much it influenced your final grade. (Mark one option per row. Please select N/A if the corresponding method was not used.) Self-assessment; Peer evaluation; Instructor evaluation; External stakeholder feedback	Multiple Choice (Grid-type): N/A + 5 item Likert-type scale
Q8	To what extent do you agree with the statement: "The assessment criteria for the capstone project were clearly communicated and understood."	Multiple Choice: 5 item Likert-type scale
	Section 3: Perceptions of Fairness and Usefulness in Assessment	
Q9	How satisfied were you with the feedback you received on your contribution to the capstone project?	Multiple Choice: 5 item Likert-type scale
Q10	Do you feel that the assessment process adequately reflected your individual contributions to the team project?	Multiple Choice: Yes; No; Somewhat
Q11	How fair do you believe each of the following evaluation methods is for judging individual contributions? (Please select N/A if this method wasn't used in your course or if you're unsure.) Peer evaluation; Self-assessment; Instructor evaluation	Multiple Choice (Grid-type): N/A + 5 item Likert-type scale
	Section 4: Use of Generative AI Tools (for projects *not* centred on Gen-AI)	
Q12	During your capstone project—excluding cases where the project itself was explicitly about Generative-AI—did you or anyone on your team make use of tools such as ChatGPT?	Multiple Choice: Yes, extensively; Yes, occasionally; No, not at all; Unsure; Prefer not to say; Not applicable—project was Gen-AI-centric

Continued on next page

Table 14 (continued)

ID	Question text	Answer format
Q13	For projects not primarily about Gen-AI, did you feel comfortable disclosing any incidental use of a Generative-AI tool to instructors or teammates?	Multiple Choice: Not applicable (I did not use AI tools); Yes, I was transparent about it; I shared it only with teammates, not with instructors; I shared it only with instructors, not with teammates; I did not disclose it to anyone
Q14	For projects that are not Gen-AI-focused, are there clear institutional guidelines on the incidental use of Generative-AI tools in capstone coursework?	Multiple Choice: Yes, clearly stated; Somewhat unclear; No, not at all; Not sure
Q15	For capstone projects whose aims were not Gen-AI-related, please indicate your agreement with the following statement: "Incidental use of Generative-AI tools compromises the fairness and integrity of capstone assessments."	Multiple Choice: 5 item Likert-type scale
	Section 5: Suggestions for Improving Capstone Assessment	
Q16	Which of the following strategies do you believe could most effectively improve the evaluation of both team and individual contributions in capstone projects? (Select all that apply)	Multiple Choice (Multi-Select): More detailed rubrics and guidelines; Regular formative feedback sessions; Incorporation of self-assessments and reflective practices; Enhanced peer evaluation processes with calibration sessions; Use of digital tools to track individual contributions; Clearer communication of assessment criteria and expectations; Other
Q17	Please share any more insight or suggestion that you have to make capstone assessment better from your point of view. What recommendations do you have to improve the experience when it comes to capstone assessment and grading?	Text Answer

Table 15: List of interview questions for instructors

ID	Question text (or interviewer-facing guidance)
	Background Questions
I1	Describe your role and experience in teaching and assessing capstone courses that are grouped based.
I2	How long have you been involved with teaching and evaluating capstone courses?
I3	Do you work with external clients or stakeholders?
	Assessment Strategies
I4	Can you briefly explain how assessment is structured in your course?
I4a	What do you assess? How much is each assessment item worth?
I4b	How often do you assess students' work (continuously, or just a single grade at the end)?
I4c	If there are external stakeholders involved, do they have any influence on the grading process (i.e. is their feedback part of the assessment)?
I5	Is there a rubric for the items assessed? Who created the rubric?
	Balance Between Individual and Collective Assessment
I6	How do you balance individual versus group contributions when grading?
I7	What challenges have you encountered in ensuring fairness when assessing group work?
I8	Do you think your assessment approach has influenced team dynamics? How?
I9	Do you allow for peer evaluations in group projects? If so, how do you incorporate them into grading?
	Assesment Challenges
I9	Have students ever raised concerns about fairness in group grading? How did you address them?
I10	How do you handle situations where one group member is not contributing equally?
I11	Have you made any changes to your group assessment approach over time? What prompted those changes?
	Survey Participation Question (Optional)
I12	Would like to participate in our survey?

Table 16: List of interview questions for Students

ID	Question text (or interviewer-facing guidance)
	Background Questions
I1	Tell me about the capstone course that you took.
I2	Did you have real-world experience before taking this capstone course (e.g., internship experience and/or work in the industry)?
I3	What was your capstone project about? Who was your client or primary stakeholder (e.g., a faculty member, an external sponsor, etc.)?
	Existing Assessment Practices
I4	How fair did you find the capstone assessment methods in evaluating your team's overall performance? (Possible probe: "What specific elements made it feel fair or unfair to you?")
I5	Did the assessment process capture your individual contributions accurately?" (Possible probe: "Can you give an example where you felt your work was—or wasn't—recognized?")
I6	Could you describe the grading strategies used to differentiate or allocate individual grades within the team? (Possible probe: "Were there rubrics, peer assessments, or instructor evaluations used?")
I7	Did you engage in peer assessments within your team? If so, how did that factor into your individual grade or feedback?
	Assessment Challenges
I8	Please describe any challenges you faced in understanding the assessment criteria used in your capstone.
I9	What challenges did you encounter with receiving feedback, such as the frequency, timeliness, or clarity of the feedback? (Possible probe: "Did you feel you had opportunities to improve based on feedback?")
I10	Have you or your teammates used generative AI tools (e.g., ChatGPT) in your capstone project? If so, what benefits or risks did you see from using them?
I11	Were there any other significant challenges, such as team dynamics or resource constraints, that affected the assessment outcomes?
	Suggested Assessment Strategies
I12	What changes in feedback practices —such as frequency, clarity, or mode of delivery— would most help you understand your strengths and areas for improvement in a team project?
I13	How could the assessment process be improved to better acknowledge both individual efforts and overall team achievements?
I14	Can collaboration platforms (e.g., GitHub, Trello) or project management tools be used to make the assessment process more transparent and fair?
I15	Are there any other changes or improvements you suggest to enhance the fairness and transparency of capstone project assessments? (Possible probe: "If you had full control over the course design, what would you definitely include or remove?")
I16	What advice would you give to future students about navigating capstone assessments effectively?
	Survey Participation Question (Optional)
I17	Would like to participate in our survey?

Table 17: List of interview questions for Stakeholders

ID	Question text (or interviewer-facing guidance)
	Background Questions
I1	Describe your role and experience as a stakeholder supporting a capstone course that was team-based?
I2	How long have you been involved as a stakeholder with capstone courses?
	Assessment
I3	Do you get involved in the assessment of the capstone? If so
I3a	What do you assess?
I3b	Do you contribute to grading directly?
I3c	Does your feedback contribute to the grading process?
I3d	How often do you provide grading input or feedback on students' work (continuously, or just a single grade at the end)?
I3e	Are you provided with a rubric for the items assessed? Are you consulted on the rubric?
	Balance Between Individual and Collective Assessment
I4	In any assessments you are involved in, how do you balance individual versus group contributions when grading or giving feedback?
I5	What challenges have you encountered in ensuring fairness when assessing group work?
I6	Do you think the external stakeholder feedback approach has influenced team dynamics? How?
	Assessment Challenges
I7	How has your experience of the assessment of capstone projects compared with your expectations?
I8	How do you think a fair assessment of capstone can be achieved?
I9	Have you ever raised concerns about fairness in group grading? How were these addressed?
I10	Have you made any changes to the way you contribute to assessment over time? What prompted those changes?
	Survey Participation Question (Optional)
I11	Would like to participate in our survey?