Student Perspectives on the Purpose of Peer Evaluation During Group Game Development Projects

Alexander Mitchell
a.mitchell@falmouth.ac.uk
Falmouth University
Cornwall, United Kingdom

Terry Greer
terry.greer@falmouth.ac.uk
Falmouth University
Cornwall, United Kingdom

Warwick New
warwick.new3@falmouth.ac.uk
Falmouth University
Cornwall, United Kingdom

Joseph Walton-Rivers
joseph.waltonrivers@falmouth.ac.uk
Falmouth University
Cornwall, United Kingdom

Matt Watkins
matt.watkins@falmouth.ac.uk
Falmouth University
Cornwall, United Kingdom

Douglas Brown
douglas.brown@falmouth.ac.uk
Falmouth University
Cornwall, United Kingdom

Michael James Scott
michael.scott@falmouth.ac.uk
Falmouth University
Cornwall, United Kingdom

ABSTRACT

Being able to work well in a team is valued in industry and beyond. As such, many university educators strive to help their students to collaborate effectively. However, it is typically the case that more than *ad-hoc* experience is needed to master teamwork. Often, students need to become reflective practitioners who learn from their experiences and enact change. Self and peer evaluation can help evoke such reflection. However, the facilitating conditions for effective learning from peer evaluation during group projects in computing are not yet well-defined. This research is an initial step in identifying these conditions. In this study, students engaged in a long-term multidisciplinary software engineering project in which they produced a digital game. They completed regular exercises in which they reflected upon and wrote about their contributions to the project as well as those of their peers. Thematic analysis of 200 responses to an open-ended question about the purpose of these exercises illustrated the student perspective: giving and receiving feedback; prompting personal reflection and improvement; supporting supervision; aiding marking; informing project planning and management; coming to a shared understanding of the status and progress of the project; exploring and reshaping group dynamics; improving project outputs; providing a system to hold group members accountable; and giving a sense of safety to raise issues without repercussion. Giving consideration to these differing perceptions will help educators to address concerns about group projects and lay the foundations for a model of effective learning from peer evaluation during student collaborations.

CCS CONCEPTS

• **Social and professional topics** → *Student assessment*; • **Applied computing** → *Collaborative learning*; • **Software and its engineering** → Programming teams.

KEYWORDS

Peer Review, Peer Assessment, Peer Evaluation, Peer Rating, Teamwork, Student Team Projects, Collaborative Learning, Group work, Software Development, Project-based Learning

1 INTRODUCTION

Employers across the computing sector need people who can work effectively in a team. Teamwork frequently appears atop surveys of what employers expect graduates to do well [19]. Rightly, student collaboration has a prominent role in computing education to fulfill this need [47]. Though, it is important to recognise that students are unlikely to become effective team workers through mere *ad-hoc* experience [27]. Instead, learning opportunities should be structured [44]. However, this can be challenging to establish, especially for those contexts still tackling exclusionary norms [20]. Peer evaluation offers one form of structure which could accommodate these challenges. It encourages students to reflect on the work of their peers alongside their own actions as practitioners; building their ability to to attain goals through critical evaluation and constructive feedback [33, 48]. They also become reflective practitioners who learn from their experiences and enact change, improving their team working competencies and advocating for positive change.

The practice of peer evaluation is well established. Though, it is important to clarify what 'peer evaluation' is. The literature is considerable, much of which can be traced back (at least) to work on sociometrics from 1940s (see [31] for a review). However, terminology varies (e.g., peer evaluation, peer rating, peer assessment, peer ranking, peer review, etc.) and the distinction between the different terms in use demands careful scrutiny. For example, Topping's [42, p.250] work explores the "arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners", largely focusing on tangible written work. Such evaluation isn't the focus of this paper. To disambiguate, the authors start with the definition provided by Kane & Lawler [23, p.555] and extend it with consideration for Boyd's [4, p.2] emphasis on the role of feedback:

> Peer [evaluation] is the process of having the members of a group judge the extent to which each of their fellow group members has exhibited specified traits, behaviors, or achievements … [in order to] provide constructive criticism and suggestions to improve weak areas and amplify strengths

To further clarify, the process sits within the context of a collaboration between people as they develop a product or service which they collectively share the responsibility for delivering. Such a tradition, commonplace in computer science education [47], encompasses the formation of *development teams* which strive to mirror their contemporaries in industry by working towards the delivery of a product or service that they themselves have created together: learning by doing. As such, this form of peer evaluation isn't as restrained to the notion of *learning teams* or *classic team-based learning* invoked in the context of other disciplines (e.g., peers reviewing each others' essays).

Following the model proposed by Salas [38], mutual performance monitoring and closed-loop communication are integral to group effectiveness. So, to give structure to such activity, students can conduct a series of exercises over the duration of a project in which they evaluate the activity of team members. A typical approach might involve assigning a rating to each team member based on specified criteria, perhaps reflecting their contributions, alongside open-ended qualitative commentary which usually yields advice on how the team and its membership can improve. These are then fed into the teaching and supervision meetings to drive improvement in a team's working practice.

Researchers continue to explore the benefits of this form of peer evaluation (see [33]). Importantly, the facilitating conditions for effective learning in the computing context are not yet well-defined. These are important to consider because some research reports domain-specific rater effects [50] which could undermine some practices. As an initial step in identifying these conditions, and because they are less established in software engineering, this paper explores the student perspective on the purpose and value of peer evaluation during group projects in higher education. As such, it addresses the following research questions: what do computing students believe the purpose of peer evaluation during group software engineering projects to be?

2 RELATED WORK

Peer evaluation has played a small but increasingly significant role in formative assessment within higher education for many years. Piaget [35] and Vygotsky [45] regard students as active, autonomous, and willing to construct knowledge from their own experience. This knowledge construction occurs within Vygotsky's social context [21], involving student-student and expert-student collaboration on real world problems or tasks that build on each person's language, skills, and experience shaped by each individual's culture [46]. Collis and Moonen [11, p.439] expand on this principle with a contribution-oriented pedagogy: "participation is not enough; the learner must also contribute to make a difference". They explain that "the most interesting contribution-oriented activities are those that are combinations of discovering and creating, comparing and discussing, and building on other learners' products" [10, p.65]. To build and effectively discuss collaborative products in higher education learners must engage in reflective practice [18], becoming reflective practitioners [15]. Though Carless and Yang highlight the need for 'workload-efficient means' of developing students' self-regulative capacities through peer-to-peer and tutor feedback [49].

Facilitating effective peer review can be onerous, but can yield efficiencies as educators become "designers and sustainers of the learning milieu; establishing conditions in which students can operate with agency" [3, p.20]. In such collaborative learning they become a 'positive model' [41] and shift the focus from staff feedback to the creation of "learning environments, the seeding of generative tasks and the fostering of interactions with and between students and staff" [3, p.20]. Such interaction requires some level of scaffolding to equip learners with the necessary communication and critical thinking skills [6].

Importantly, though, peer evaluation itself can help to "develop [these] important high-order (generic) skills such as critical evaluation and communication, [… and] self-assessment" [32, p.672]. Other key benefits observed during peer evaluation are reciprocity and dependency [37]: the student identifies as part of a wider group and its associated benefits. An effective collaborative learning environment for higher education also requires that students look to each other as knowledge providers and develop independent frameworks of enquiry, challenging the convention of regarding the teacher as the sole source of information [25]. Carless and Lui assert that peer evaluation is an end in itself [28], arguing that it equips students with diverse skills. Skills that are valued in industry like: critical reflection [30]; listening to and acting on feedback; sensitively assessing; and providing feedback on the work of others. These skills are also a part of industry practice itself, seeking feedback and self-evaluating being part of production, particularly when following agile approaches [16].

Taking the game development industry as an example, by nature it is "iterative, experiential, and inter-disciplinary, understanding peer feedback is key to improving the game creation process as a whole" [39, p.427]. Though it has become a popular context in which to support computing education [9], research into the use of peer evaluation during game development projects is less developed in the literature with only one prominent study in 2011. This study found an instructor-led peer evaluation

process during a group game design project lead to higher student satisfaction and perceptibly knowledgeable engagements between the instructor their cohort, but the scores relating specifically to peer review were more neutral and therefore inconclusive [40].

Some reticence on the part of students may be due to challenges observed in numerous studies. For some it is lack of confidence in the process [8] feeling unqualified to assess fellow students, for others it is the problem of the 'free rider' [26, 34, 43] where low participation students in collaborative projects benefit from the work of their peers. There is also a considerable issue associated with biased marking based on affiliation within a group [7]. Stress is cited as an issue, stemming from a dilemma of truthful and objective marking versus the loyalty owed to group members [36]. Perceptions of bias can vary based on the subtle differences between peer review, peer ratings, and peer rankings [23]. There is also some evidence to suggest that students can behave in anti-reciprocal ways. For example, it has been observed that some students start to write shorter (worse) reviews when they receive longer (better) evaluations [24]. Ultimately any framework of peer assessment needs to consider the "emotional responses and psychological needs" [49, p.292] of students. Notably, under what circumstances are critical comments valued and appreciated, and acted upon to yield positive change and better performance.

The identification of tools and practices which form an environment in which peer evaluation can flourish is an active area of research. Recently, one institution in the UK investigated the inclusion of peer review in group work, striving for a balance between group and peer assessed marks [17]. They drew upon approaches spanning oversight, process, and system design, including frameworks to improve the way in which students are held accountable in the way marks are awarded (i.e., [28]). To mitigate the 'free rider' problem, Lu and Tu [43] suggest a ranking system applied by students based on contribution. Points are awarded for engagement, conversely points are deducted for untruthful behaviour. This approach is further developed in a study that emphasises the development of 'social competencies.' Students can far better assess the criteria of team work; organizing and coordinating abilities than instructors, but this requires student literacy in feedback [3] and systems to analyse frequency and message sentiments in text-based communications within collaborative groups [26]. Indicators of team success through 'interaction analysis' is also emerging as a method to "overcome bias emerging from interdisciplinary, cognitive and meta cognitive differences in participants" [13, p.289].

Though software is being developed to help overcome such challenges (e.g., wiki tracking [44], plug-ins for virtual learning environments [2]), there are gaps in the literature on what peers assess and how, which are needed to inform the design of tools shaping peer evaluation in the computing context, especially during game projects. There is an implied need to capture productivity outcomes alongside the softer skills associated with negotiating personalities and behaviours to sustain group cohesion and effective collaboration. Notably, a thirst for more 'socially translucent systems' [14, p.62] which promote visibility, awareness, and accountability. Therefore, taking a student-centred approach to exploring the purpose and value of peer evaluation is a ripe opportunity for research.

3 CONTEXT

The primary research question which this paper explores concerns the student perspective of the purpose of peer evaluation. Though it is likely that the findings can inform other contexts, it is important to acknowledge the domain-specificity highlighted in the literature review. Due to variation in practice across the sector, the nature of the question in light of this, and the games focus of the particular context studied, insight into the context in which this study has been conducted is provided here.

At the authors' institution, peer evaluation is practiced as part of multidisciplinary team projects on software engineering modules. These include students across a range of undergraduate awards in fields aligned to the UK's digital economy including computing, digital art, design, production, audio technology and writing. Though, typically focus on making games using the Unity and Unreal engines. Simulating the games industry is a primary aim, with students working in groups formed of peers on different courses. Any given student will work in several unique teams during their time in the department, at least once at each level of study (i.e., Levels 4–7 in the FHEQ[1]), typically seeing the process of game creation from concept through to a complete product across a study period of 26 weeks. This follows a history of embedded teamwork, established group work strategy praised by external accrediting bodies and wide disciplinary range has functioned as a testing ground for many initiatives in this area since it was founded in 2014, gaining significant international recognition for its methods and was one of the case studies in. Being central to the ethos of the department, peer evaluation and its link to learning has become central to its group work strategy.

From the staff perspective, beyond being a learning aid, peer evaluation needed to serve multiple agendas: to enable students to discuss, sometimes to vent about their working experience with their peers; to enable formal assessment against set learning outcomes by supervisors and module leaders; and to fulfil its potential to improve the actual outputs of the group projects. Notably, by addressing weaknesses in production methods in addition to promoting critique and improvement of contributions to the executable build (i.e., code, assets, user experience, etc.). The peer evaluation exercises themselves were conducted in regular team supervision meetings facilitated by online forms to allow anonymisation of student comments, with the supervisor delivering the feedback verbally to the group and facilitating a retrospective discussion between group members.

The participants were 200 university students enrolled on these modules, representing a response rate of ~40%. Of these, 35% were in their first stage of study, 34% the second, 26% in the third and final undergraduate stage, and 4.5% on taught postgraduate courses. Predominantly, they identified as male 73% with only 25% identifying as female and only 2% non-binary. The mean age was 21.4 years, with 71% representing the 18–21 age bracket and the remainder in the 22–43 range. This broadly corresponds to known population statistics for the department. Typical entry requirements for 112–120 UCAS[2] points with roughly half the entrants having previous software development experience.

---

[1]Framework for Higher Education Qualifications
[2]Universities and College Admissions System

4 METHODOLOGY

This study was conducted across one postgraduate and three undergraduate modules on practical software engineering; all with learning outcomes about teamwork, all with assessments delivered collectively as a team, and all tailored to the development of digital products and services related to the games industry. Each of these modules followed the same general structure of timetabled learning activities and used the department's group working strategy placing an emphasis on peer evaluation.

Following the conclusion of these modules in May 2021, students were invited to complete a questionnaire about their experience of peer evaluations. The questionnaire was conducted using Microsoft Forms, and hosted within our Microsoft Azure tenancy. The survey was promoted by email, by livesteam at the departmental ask-me-anything, as part of our annual Expo, and follow-up email invitations for those students who had not yet responded. As an incentive to participate, there was a prize draw for those completing the survey. The questionnaire itself gathered demographic information and students were asked to provide plain-text responses to two open-ended questions: "briefly state what you believe the purpose of peer evaluation in group project teams is"; and "Please comment on anything else that comes to mind about peer evaluation and/or group projects'.

The responses were then analysed using the six-stage thematic analysis procedure proposed by Braun and Clarke [5], with some adaptation to the later stages and an extra stage of standardisation so re-coding could occur under a unified schema and quantised for counting (i.e., to determine prevalence). This process started with three coders independently reviewing the data to identify themes: a pattern of meaning shared across instances that embody some common concept or idea. The raters independently determined what themes they could identify within the data, initially following an inductive coding approach. After this initial pass through the first five stages of thematic analysis, up to where each rater labelled the themes they had identified, the raters reconvened with all authors to discuss their findings and refine a common set of codes (i.e., a schema). Once this was agreed, the raters again re-coded the data according to this new schema. Once a common set of codes was established the coders revisited the data rating each student response against the appropriate codes. Then, a process of standardisation was followed which involved comparing all the codes and identifying discrepancies in the way they had been deductively applied to the data. Once this stage had been completed there was a final review of the data set to highlight any points of disagreement between the coders, and the remaining authors were consulted to break stalemates.

This process of consensus coding was repeated until agreement exceeded 90% responses, giving a deeper understanding of how each of the themes was defined, how broad its meaning was, and to distinguish between the themes in complex edge cases. It also allowed for the analyse of the data to be refined beyond a single pass and ensured that the information was correctly coded reducing clerical errors during the coding process. Once this final pass was completed for the themes, it was possible to determine the relative proportion each theme emerged in the underlying data, with some consideration for the level of concordance between the raters.

5 RESULTS & ANALYSIS

5.1 Purpose of Peer Evaluation

A word cloud of the student responses (following stop word removal and lemmatization) can be found in Figure 2. The thematic analysis involved three raters who reviewed the data independently to each other. Three rounds of independent analysis and discussion took place to establish a schema agreed between all of the raters and the remaining authors. A further two iterations of re-coding were needed to converge on consensus, reaching an overall agreement level 92.3%. Nine key themes were identified: (i) giving and receiving feedback; (ii) prompting personal reflection and improvement; (iii) supporting supervision; (iv) aiding marking; (v) informing project planning and management; (vi) exploring and reshaping group dynamics; (vii) improving project output; (viii) providing a system to hold group members accountable; and (ix) giving a sense of safety to raise issues without repercussion. Within these themes, agreement varied between 82.5% and 99.7%, with some themes having complex characteristics suggestive of sub-themes. The Krippendorff's alpha statistic [22] has been calculated using SPSS `20.0.0` and provided for each individual theme to help judge measurement reliability. The frequency with which these themes occurred in the data is illustrated in Figure 1, showing that (i), (ii), and (vi) were the most prevalent amongst the students surveyed. The most complex themes were (v), (vi), (vii) and (viii) so their reported prevalence should be interpreted with caution.

5.1.1 Giving and Receiving Feedback. This theme relates to giving and receiving advice, reference to constructive criticism or advice, and to the barriers associated with feedback such as the awkwardness of giving effective feedback. This theme was observed in 52.5% of cases (α = .790).

5.1.2 Prompting Personal Reflection and Improvement. This theme spans reference to self-improvement, the driving of peer improvement, as well as the identification of strengths and weaknesses as well as other opportunities to improve. Overall it was observed in 46% of cases (α = .703).

5.1.3 Supporting Supervision. The idea that the peer evaluation was used to inform the supervisor of the project, given they were not aware of what was occurring internally within the team. Typically, references included giving opportunities to better understand how the team was working, what support they needed from their supervisor, or highlighting issues that they needed advice with. This theme was observed in 20.5% of cases (α = .807).

5.1.4 Aiding Marking. Using the evaluations provided by members of the team to help the course team with assessment activity. Notably, tracking what people are contributing. Few students perceived peer evaluation as a way to assist in the marking of their contribution to group (11% of cases, α = .852).

5.1.5 Informing Project Planning and Management. This theme encompassed a variety of benefits that helps a team stay on track with their projects, notably criticising the project management, plan, workflows, or compliance with agile methodologies. This theme only appeared in 8% of cases (α = .581).
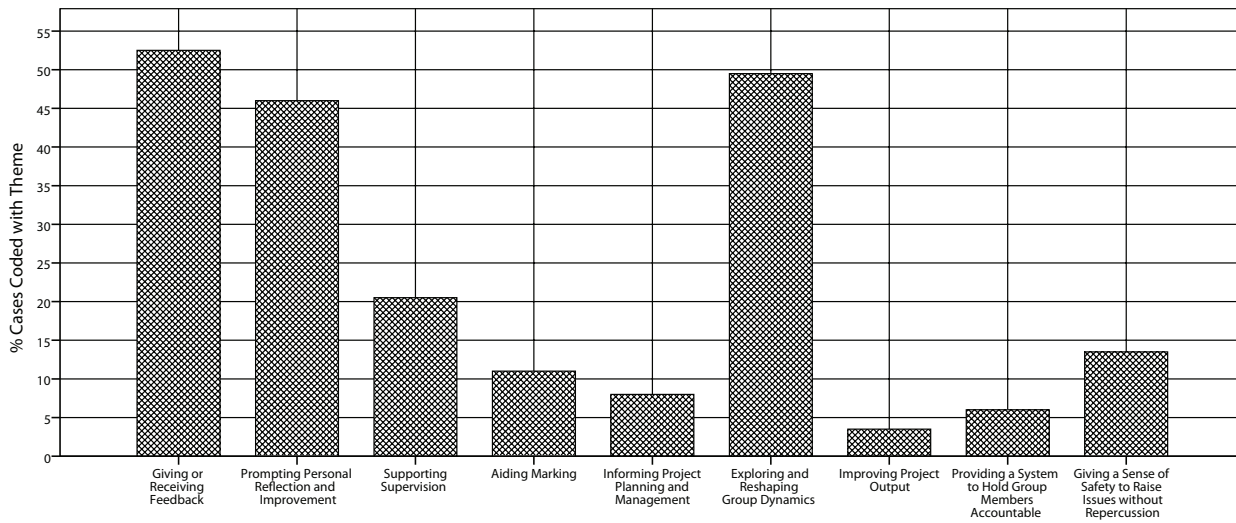
Figure 1: Bar Chart Illustrating the Proportion of Students Endorsing Each of the Identified Themes

Figure 2: Word Cloud of Terms Reflecting the Student Perspective on Peer Evaluation

5.1.6 Exploring and Reshaping Group Dynamics. Many cases (49.5%, $\alpha = .538$) suggested the purpose of peer evaluation was to facilitate in understanding their team, typically by: recognising teammates contributions to the project, improve the well-being of the team, and aiding conflict resolution. Anxieties associated with team functioning and working practice as well as the mutual trust needed for honest evaluations were also captured by this theme.

5.1.7 Improving Project Output. This theme addressed the improvement of the actual output of the project—the work produced by the team, usually during the sprint preceding the evaluation. Typically, referencing specific critique of deliverables under version control, or ways to improve the state of the executable build. This theme was observed in 3.5% of cases ($\alpha = .610$).

5.1.8 Providing a System to Hold Group Members Accountable. A small number of cases (6%, $\alpha = .543$) emphasised the role of peer evaluation as an incentive to encourage peers to contribute to the project, or in identifying students that did not pull their weight.

5.1.9 Giving a Sense of Safety to Raise Issues Without Repercussion. This referred to the statements about peer evaluation being a safe space to raise concerns without worry or fear of consequences, or creating unnecessary friction with the rest of the team. This theme was observed in 13.5% of cases ($\alpha = .762$).

5.2 Other Comments

When prompted with the optional question asking for additional comments, 115 of the 200 participants provided a response (ignoring 'n/a' responses). Among the responses for this question, the key themes identified were: (i) inconsistency in supervisor feedback; (ii) onerousness of the process; (iii) clarity of information about peer evaluation; (iv) appropriate etiquette in peer evaluation; (v) desire for feedback clarity; (vi) concerns of bias; and (vii) insight into the student experience.

5.2.1 Inconsistency in Supervisor Feedback. This theme was focused around the different styles that lecturers used to convey the feedback to the students. Students noted that different supervisors take different approaches and were unhappy with these inconsistencies.

5.2.2 Onerousness of the Process. Students noted that the feedback forms took a long time to fill in, and had a desire to see the process streamlined so that it took less time to complete. Common reasons for this was the amount of feedback written, and that large team sizes making filling in the forms take a long time.

5.2.3 Information Clarity. There was a perceived lack of transparency in how peer evaluation data was being used. This was often expressed as a desire to have the data made more available to the team and also that it should be used to provide accountability for team members behaviour and contributions.

5.2.4 Apparent Etiquette in Peer Evaluation. Respondents reported that they did not want to appear overly negative in feedback, and would avoid bringing up issues. There were also mentions of a desire

to not want to offend students by providing negative feedback (and that doing so could cause conflict within the team).

5.2.5 Desire for Feedback Clarity. Students expressed uncertainty about the uses of the data collected from peer evaluation and how it was being used to measure their contribution.

5.2.6 Concerns of Bias in Feedback Process. This theme dealt with student attitudes towards peer evaluation marks. Students noted that students which were more vocal or popular may be perceived as getting better marks in peer evaluations. Students also noted that they would prefer that staff did not use peer evaluation scores when determining student grades. This theme also featured students reporting that they felt that students that did not turn up were not punished harshly enough. It is worth noting there is a high proportion of allusions to accountability (~20%) here.

5.2.7 Insight into the Student Experience. Reporting sentiments, both negative and positive, that largely reinforced the themes presented in the previous section. For example, endorsing the anonymous nature of the evaluation to enable feedback to be expressed without repercussions. There were also comments that reflected on their own experience being enjoyable or negative, but remarking that they weren't sure whether their experience was shared across the whole cohort.

6 DISCUSSION

A broad range of purposes have been attributed to peer evaluation in the context of group projects. Most trended primarily about giving and receiving feedback, alongside the strongly related theme of personal reflection. This accords with the development of critical reflective skills being a reported focus of peer evaluation in the literature [32]. Though not as strong a theme, exploring the teams dynamics also had considerable presence. It seems that students perceive peer evaluation as an opportunity to improve the team working experience and respond to weaknesses they perceive in the way their team is tackling their assigned project.

There were, however, several notions that the authors anticipated which were notable in their absence. Despite the prominence of exploring roles within the team, there were few responses suggesting whether peer evaluation was a useful measure of how the project was progressing or reflecting on the broader impact the group's chosen software engineering methods. It wasn't, for example, perceived as a mechanism to help manage the 'scope' of their project nor to tackle other difficult challenges in software project management like estimation [1].

Focal points instead tended to highlight interpersonal frictions and differences between the disciplines involved—largely, due to lack of organisation or communication. It was surprising not to see any mention of whether or not contributions were appropriately aligned to the needs of the project or of the team, or concerns about 'presenteeism' [12, 29]. Even accountability didn't appear to trend particularly strongly when prompted about purpose.

However, in the further comments there was strong sentiment regarding accountability, for both attendance and performance. This is consistent with reports in the literature on the desire to ensure that 'free riders' are not awarded marks unfairly [26, 34, 43]. So, while it may not currently be something the students perceives

as a reason or benefit of peer evaluation, there is at least some desire to see peer evaluation leveraged a way to stimulate engagement.

Despite this, and aiding marking being endorsed by a few as a purpose, there is little desire to formally link peer evaluation with marking. Again, further comments illustrated strong views to this end. There is, perhaps, a lack of mutual trust between members of these students teams. Bias and popularity contests are one concern [23]. Though some students reported they felt they couldn't be honest with their feedback as even when given the ability to give feedback privately they felt uncomfortable providing negative feedback to their peers. There is evidence of reluctance to negatively rate their peers such as the close proximity in which they work with their peers, supporting the findings of Pond [36]. It's worth noting that some students reluctance to provide negative feedback could be the direct cause of the lower number of students saying the peer evaluation activity lead to conflict within their groups. They felt compelled not to offend or otherwise feared peers lacked the maturity to handle critique. This also might be related to the lack of confidence in their own reflective skills, as observed elsewhere [8].

The data suggests that students can have quite different experiences with peer evaluation. This could depend on the supervisor, but could also be a function of other factors such as who they were grouped with, the culture that emerges in a team, trust between members, etiquette, onerousness, consistency of practice, and others. Some students report that such challenges devalue the process. Others claimed that although they had a positive student experience, they knew someone who hadn't. This inconsistency highlights a potential shortcoming of peer evaluation and suggests more work is needed to model the conditions with which peer evaluation is effective.

7 CONCLUSION

This paper addresses the absence of case studies on peer evaluation in games development contexts. It endorses a student-centred approach to exploring peer evaluation, answering the question: what do computing students believe the purpose of peer evaluation during group software engineering projects is? It has provided insight that will be of interest to computing educators. In exploring the student perspective, nine themes have emerged, highlighting a variety of perceptions on what value peer evaluation offers. This data suggests that students have differing expectations of peer evaluation, many of which are not mutually exclusive but are deserving of further recognition by educators. Further analysis of open-ended comments suggests areas of concern with some students seemingly having vastly different experiences of peer evaluation, suggesting directions for future work. The study has several limitations. Notably, the authors acknowledge this data is limited to a single case study at one institution. It is also based solely on qualitative data. However, the study has set the stage for larger studies to be carried out, exploring how these different conceptions of peer evaluation can be accommodated while still delivering a meaningful opportunity for students to engage with these processes for mutually enhanced learning of group work. The nine themes will also form the foundation for a future model of effectiveness which can be used to define the facilitating conditions under which learning of teamwork will improve.

REFERENCES

[1] 2014. Teaching Software Project Management. In *Overcoming Challenges in Software Engineering Education: Delivering Non-Technical Knowledge and Skills: Delivering Non-Technical Knowledge and Skills*, Liguo Yu (Ed.). IGI Global.

[2] Gabriel Badea, Elvira Popescu, Andrea Sterbini, and Marco Temperini. 2019. Integrating enhanced peer assessment features in moodle learning management system. In *Foundations and Trends in Smart Learning*. Springer, Singapore, 135–144.

[3] David Boud and Elizabeth Molloy. 2012. Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education - ASSESS EVAL HIGH EDUC* 38 (Jan. 2012), 1–15. https://doi.org/10.1080/02602938.2012.691462

[4] Ronald TC Boyd. 1989. Improving teacher evaluations. *Practical Assessment, Research, and Evaluation* 1, 1 (1989), 7.

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

[6] David Carless. 2019. Feedback loops and the longer-term: towards feedback spirals. *Assessment & Evaluation in Higher Education* 44, 5 (July 2019), 705–714. https://doi.org/10.1080/02602938.2018.1531108

[7] Christina M. Cestone, Ruth E. Levine, and Derek R. Lane. 2008. Peer assessment and evaluation in team-based learning. *New Directions for Teaching and Learning* 2008, 116 (2008), 69–78. https://doi.org/10.1002/tl.334 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/tl.334.

[8] Winnie Cheng and Martin Warren. 1997. Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education* 22 (Jan. 1997), 233–239.

[9] Bill Clark, Jerry Rosenberg, Terrel Smith, Stu Steiner, Scott Wallace, and Genevieve Orr. 2007. Game development courses in the computer science curriculum. *Journal of Computing Sciences in Colleges* 23, 2 (2007), 65–66.

[10] B. Collis and J. Moonen. 2006. The contributing student: learners as co-developers of learning resources for reuse in Web environments.

[11] Betty Collis and Jef Moonen. 2009. Contribution-oriented pedagogy. In *Encyclopedia of Distance Learning, Second Edition*. IGI Global, 439–446.

[12] Cary L Cooper. 1998. The changing nature of work [1]. *Community, Work & Family* 1, 3 (1998), 313–317.

[13] Thanasis Daradoumis, Fatos Xhafa, and Angel Juan. 2006. A Framework for Assessing Self, Peer, and Group Performance in E-Learning. 279–294. https://doi.org/10.4018/9781591409656.ch012

[14] Thomas Erickson and Wendy Kellogg. 2000. Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Trans. Comput.-Hum. Interact.* 7 (March 2000), 59–83. https://doi.org/10.1145/344949.345004

[15] Nancy Falchikov and David Boud. 1989. Student self-assessment in higher education: A meta-analysis. *Review of educational research* 59, 4 (1989), 395–430.

[16] Frederik M Fowler. 2019. The Sprint Retrospective. In *Navigating Hybrid Scrum Environments*. Springer, 97–100.

[17] Souto Garcia, C Striolo, and M Vogel. 2017. Peer assessing individual contributions in a group project. https://discovery.ucl.ac.uk/id/eprint/10063005/1/SRHE%20roundtable_Pilar_v5c.pdf

[18] Graham Gibbs. 1988. Learning by doing: A guide to teaching and learning methods. *Further Education Unit* (1988).

[19] Wouter Groeneveld, joost Vennekens, and kris Aerts. 2019. Software Engineering Education Beyond the Technical: A Systematic Literature Review. *Proceedings of the 47th SEFI Conference 2019* 47, 1607–1622. https://www.sefi.be/proceedings/

[20] Alison Harvey. 2019. Becoming Gamesworkers: Diversity, Higher Education, and the Future of the Game Industry. *Television & New Media* 20, 8, 756–766.

[21] Samuel J. Hausfather. 1996. Vygotsky and Schooling: Creating a Social Context for Learning. *Action in Teacher Education* 18, 2 (1996), 1–10. https://doi.org/10.1080/01626620.1996.10462828 arXiv:https://doi.org/10.1080/01626620.1996.10462828

[22] Andrew F Hayes and Klaus Krippendorff. 2007. Answering the call for a standard reliability measure for coding data. *Communication methods and measures* 1, 1 (2007), 77–89.

[23] Jeffrey S Kane and Edward E Lawler. 1978. Methods of peer assessment. *Psychological bulletin* 85, 3 (1978), 555.

[24] Yasmine Kotturi, Andrew Du, Scott Klemmer, and Chinmay Kulkarni. 2017. Long-Term Peer Reviewing Effort is Anti-Reciprocal. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. ACM, Cambridge Massachusetts USA, 279–282. https://doi.org/10.1145/3051457.3054004

[25] Ilana Lavy and Aharon Yadin. [n.d.]. Team-Based Peer Review as a Form of Formative Assessment - The Case of a Systems Analysis and Design Workshop. 21 ([n.d.]), 15.

[26] Hye-Jung Lee and C. Lim. 2012. Peer Evaluation in Blended Team Project-Based Learning: What Do Students Find Important? *Educational Technology and Society* 15 (Jan. 2012), 214–224.

[27] R. Lingard and S. Barkataki. 2011. Teaching teamwork in engineering and computer science. In *41st IEEE Frontiers in Education Conference*. https://doi.org/10.1109/FIE.2011.6143000

[28] Ngar Liu and David Carless. 2006. Peer feedback: The learning element of peer assessment. *Teaching in Higher Education - TEACH HIGH EDUC* 11 (July 2006), 279–290. https://doi.org/10.1080/13562510600680582

[29] Bruce Macfarlane. 2015. Student performativity in higher education: Converting learning as a private space into a public performance. *Higher Education Research & Development* 34, 2 (2015), 338–350.

[30] Catherine Moore and Susan Teather. [n.d.]. Engaging students in peer review: Feedback as learning. ([n.d.]), 16.

[31] Jane Srygley Mouton, Robert R Blake, and Benjamin Fruchter. 1955. The validity of sociometric responses. *Sociometry* 18, 3 (1955), 181–206.

[32] Raoul Mulder, Chi Baik, Ryan Naylor, and Jon Pearce. 2014. How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education* 39 (Aug. 2014). https://doi.org/10.1080/02602938.2013.860421

[33] David Nicol, Avril Thomson, and Caroline Breslin. 2014. Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education* 39, 1 (Jan. 2014), 102–122. https://doi.org/10.1080/02602938.2013.795518

[34] Elizabeth Pfaff and Patricia Huddleston. 2003. Does It Matter if I Hate Teamwork? What Impacts Student Attitudes toward Teamwork. *Journal of Marketing Education - J Market Educ* 25 (April 2003), 37–45. https://doi.org/10.1177/0273475302250571

[35] Jean Piaget. 2003. *The psychology of intelligence*. Routledge.

[36] Keith Pond, Dave Coates, and Ofelia Palermo. 2007. Student Experiences of Peer Review Marking of Team Projects. *International Journal of Management Education* 6 (Oct. 2007). https://doi.org/10.3794/ijme.62.190

[37] Krishneel Krishna Reddy. [n.d.]. Improving Students' Assessment and Evaluation Experience in Higher Education: A Formative Peer Review Perspective. ([n.d.]), 231.

[38] Eduardo Salas, Dana E Sims, and C Shawn Burke. 2005. Is there a "big five" in teamwork? *Small group research* 36, 5 (2005), 555–599.

[39] Joseph Seering, Ray Mayol, Erik Harpstead, Tianying Chen, Amy Cook, and Jessica Hammer. 2019. Peer Feedback Processes in the Game Industry. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Barcelona Spain, 427–438. https://doi.org/10.1145/3311350.3347176

[40] Amber Settle, Charles Wilcox, and Chad Settle. 2011. *Engaging game design students using peer evaluation*. https://doi.org/10.1145/2047594.2047614 Journal Abbreviation: SIGITE'11 - Proceedings of the 2011 ACM Special Interest Group for Information Technology Education Conference Pages: 78 Publication Title: SIGITE'11 - Proceedings of the 2011 ACM Special Interest Group for Information Technology Education Conference.

[41] Junko Shimazoe and Howard Aldrich. 2010. Group Work Can Be Gratifying: Understanding & Overcoming Resistance to Cooperative Learning. *College Teaching* 58 (March 2010). https://doi.org/10.1080/87567550903418594

[42] Keith Topping. 1998. Peer assessment between students in colleges and universities. *Review of educational Research* 68, 3 (1998), 249–276.

[43] Yanbin Tu and Min Lu. [n.d.]. Peer-and-Self Assessment to Reveal the Ranking of Each Individual's Contribution to a Group Project. ([n.d.]), 10.

[44] Rebecca Vivian, Katrina Falkner, Nickolas Falkner, and Hamid Tarmazdi. 2016. A Method to Analyze Computer Science Students' Teamwork in Online Collaborative Learning Environments. *Trans. Comput. Educ.* 16, 2, Article 7 (Feb. 2016), 28 pages. https://doi.org/10.1145/2793507

[45] Lev Semenovich Vygotsky. 1980. *Mind in society: The development of higher psychological processes*. Harvard university press.

[46] Lev S Vygotsky. 2012. *Thought and language*. MIT press.

[47] Mattias Wiggberg. 2008. *Unwinding processes in computer science student projects*. Ph.D. Dissertation. Uppsala universitet.

[48] Keith Willey and Anne Gardner. 2010. Investigating the capacity of self and peer assessment activities to engage students and promote learning. *European Journal of Engineering Education* 35, 4 (Aug. 2010), 429–443. https://doi.org/10.1080/03043797.2010.490577

[49] Min Yang and David Carless. 2013. The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education* 18 (April 2013). https://doi.org/10.1080/13562517.2012.719154

[50] Bo Zhang, Lucy Johnston, and Gulsen Bagci Kilic. 2008. Assessing the reliability of self-and peer rating in student group work. *Assessment & Evaluation in Higher Education* 33, 3 (2008), 329–340.