An Exploratory Analysis of Student Experiences with Peer Evaluation in Group Game Development Projects

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

Michael James Scott
michael.scott@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

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

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

ABSTRACT

Collaborative projects are commonplace in computing education. They typically enable students to gain experience building software in teams, equipping them with the teamwork skills they need to be competitive in the labour market. However, students often need encouragement to reflect upon and synthesise their experience to attain the most learning. Peer evaluation offers one such approach, but the conditions which facilitate effective peer evaluation have not yet been established. This paper seeks to provide insight into student experiences with peer evaluation. It builds upon prior qualitative work, analysing quantitative data collected through a questionnaire taken by undergraduate students on a collaborate digital game development module. An exploratory factor analysis identifies seven dimensions of variance in the student experience: perceived impact; arbitrary influence; inconsistency; team cohesiveness; assessment pressure; ease and professionalism. Correlation analysis suggests some factors such as arbitrary influence, team cohesion, assessment pressure, and professionalism are associated with attained learning, whilst factors such as inconsistency and onerousness are not. This informs the development of a conceptual framework, suggesting focuses which facilitate effective peer evaluation. Expanding this conceptual framework and validating it across different demographics, contexts, and project types are suggested as avenues for further investigation.

CCS CONCEPTS

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

KEYWORDS

Peer, Evaluation, Review, Assessment, Rating, Collaboration, Software Development, Project-based Learning

This work is licensed under a Creative Commons Attribution 4.0 International License.
UKICER2022, September 1–2, 2022, Dublin, Ireland
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9742-1/22/09.
https://doi.org/10.1145/3555009.3555021

ACM Reference Format:
Alexander Mitchell, Michael James Scott, Joseph Walton-Rivers, Matt Watkins, Warwick New, and Douglas Brown. 2022. An Exploratory Analysis of Student Experiences with Peer Evaluation in Group Game Development Projects. In *The United Kingdom and Ireland Computing Education Research (UKICER) Conference (UKICER2022), September 1–2, 2022, Dublin, Ireland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3555009.3555021

1 INTRODUCTION

Employers in many sectors seek computing graduates who can collaborate effectively, with interpersonal and communication skills being highly valued [18]. Team projects in which students deliver digital products or services following methods that simulate industry can provide students with an opportunity to develop such skills [34]. However, it would be naive to assume that mere exposure to such an opportunity as part of a university course would be sufficient [27]. Rather, to help students acquire these skills, some form of structure is needed to direct and aid learning [49]. Encouraging students to reflect upon and synthesise their experience is one such approach, and can be achieved through peer evaluation. This enables students to become reflective practitioners who critically engage with their teams.

This approach to supporting team-work is highlighted as beneficial by Clark [9]. Though, despite the prominence of group projects in computing education [50], there is a lack of evidence on the conditions which facilitate effective peer evaluation, particularly in the game development context. This paper examines the experience of students with peer evaluation on a large group game development module. Understanding student experiences of peer evaluation will provide insight into what helped or hindered their teamwork, thus informing the design of interventions and tools that will ultimately improve attainment.

2 RELATED WORK

There is considerable literature on peer evaluation (e.g., [21, 30, 43, 45, 48]), though much of the terminology varies (e.g., peer evaluation, peer review, peer assessment, peer rating, peer ranking, etc.) and there are subtle distinctions between them. It is typically a formative process in which students constructively evaluate one another's work during group projects [11, 36]. This is similar to peer feedback, which also describes a formative feedback

process, but is instead usually framed in terms of a dialogue [29, 53]. When discussing summative processes, the term 'peer assessment' is also often used. This can refer to either the work, performance or both [13, 45]. In this paper, the authors define peer evaluation based on the work of Kane & Lawler [21, p.555] and then expand upon it through the lens of Boyd's [5, p.2] work on 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 ... [to] provide constructive criticism and suggestions to improve weak areas and amplify strengths.

This definition places emphasis on reflective practice [17], and on the journey to becoming a reflective practitioner [14]. This has utility in group project contexts, for example in supporting students to develop an effective team. Consider for example Tuckman's [47] five-stage model of team development: forming, storming, norming, performing and adjourning. Teams trapped in the 'storming' stage will experience disruptions which inhibit their productivity. However, enacting steps from shared reflective insight can help a team overcome this 'storming' stage.

Reflection can also help drive and inform the process of self-transformation [29], encouraging students to develop their own methods of enquiry, rather than looking to educators as the single source of knowledge [24]. Such an approach helps students to develop ways of receiving, responding to, and acting on feedback, which are among skills valued in industry [33].

The literature describes many tools which can be used to facilitate peer evaluation, with varying emphases on the review of group member contributions or of work, as well as on the sharing of peer ratings or of peer feedback [2, 12, 30, 49]. A common thread in this work is the desire for 'workload-efficient means' of supervising group work [25]. Boud [3], Race [39], alongside Gärdebo and Wiggberg [16] suggest that peers offer largely untapped insights into teamwork which could assist with feedback at scale. In addition, although often not their primary aim, this helps minimise staff assessment workload by reducing administration. Beyond the promise of increased productivity, the foci of research into the design of such tools is varied.

Prominent focus is placed upon peer rating due to the existence of 'free riders' and 'saboteurs': students who have engaged poorly with a group endeavour but still seek to attain high marks through the efforts of their peers [28, 36, 46]. To address this challenge, Tu and Lu [46] propose the use of a ranking system that recognises contribution and highlights dishonest behaviour. Several systems (e.g., [51]) implement similar approaches. Rating and its relationship to marking is an area which evokes much concern. There are concerns that students mark based on social standing rather than rigorous application of assessment criteria [6], and that other biases lead students to "give themselves the highest amount both too often and not enough" [19, p.9]. Many other forms of bias also complicate the process [9, 10, 42, 44].

Peer evaluation strives to mitigate these biases and drive constructive behaviours by placing emphasis on formative feedback. Fellenz [15] reports that concerns regarding the influence of friendship or reciprocity on feedback in peer evaluations is ill founded as the literature suggests these effects are not significant. However, even peer evaluation is vulnerable to bias and other concerns. Many students report they lack confidence in their abilities to adequately evaluate their peers [7] and thus feel unable to articulate meaningful feedback, or report they experience stress as they balance being truthful and professional with social pressure to avoid anything that might compromise good working relationships [37]. Though it exists, the discourse on 'feedback literacy' [4, 52] and how to improve it is deserving of more attention.

To address these gaps, student perspectives would seem an appropriate place to start. Multiple studies show positive sentiment [35, 36] and effectiveness [52], particularly where peer evaluation is transparent [1]. However, such studies provide only limited insight into the facilitating conditions under which peer evaluation is effective. A mature conceptual framework to examine these conditions has yet to emerge. Furthermore, there are comparatively few studies in computing contexts, or in game development contexts specifically. The few that do exist show that industry-inspired approaches to peer evaluation can improve student satisfaction, but are inconclusive with respect to any link to learning outcomes [41]. Therefore, the way in which the different facets of peer evaluation relate to attainment is unclear. Given the popularity of using game development as a context for computing education [8], and especially so with game developers valuing iterative peer feedback throughout their production processes [40], further investigation is warranted.

3 RESEARCH CONTEXT

This paper seeks to address the following research questions:

- RQ1. From the student perspective, in which key ways does their experience of peer evaluation vary?
- RQ2. To what extent do these factors correlate with effort, attainment, and with each other?

These are investigated in the context of a group game development module in the Games Academy at Falmouth University. This module involves making a digital game, starting with an initial concept and taking it through to a complete product, across a 20-week period. The projects are multidisciplinary in nature, as the module is shared across different courses. Teams are composed of people studying a range of fields aligned to the digital economy, with many which are conventionally associated with computing but several others which are adjacent to it. This particular module runs in the first stage of each programme, at level four according to the Framework for Higher Education Qualifications [38].

There is a group working strategy applied to all collaborative modules across the department. There are typically 8-12 students in each group, and each is supervised by an academic. Group allocation methods are permitted to vary, but for this particular module it was random, though the module leader did sanity check the allocation for aberrations (e.g., disciplinary mix, demographic diversity, etc.). The strategy describes how peer evaluation is conducted, which involves each student reporting to their supervisor through an online form on a fortnightly basis. This is

facilitated through the Moodle[1] virtual learning environment using the Feedback Fruits[2] third-party plugin. This is a tool which manages the collection, storage, and presentation of group member evaluations. They write comments for each of their peers in the team. Comments are anonymous for students, but not supervisors. Its form provides a field for constructive feedback, which is shared with the other team members, alongside fields for confidential notes for only the supervisor to see. This primarily informs discussion and facilitation in regular timetabled supervision meetings. There is an indirect connection to assessment, since markers are aware of the feedback, alongside other indicators of attainment including supervisory notes, version control logs, digital task boards, and observed meetings. Grades are awarded to each student by combining team output (i.e., the game) at the end of the project with their individual contributions, value-added, and agile practice.

There were 133 students participating in the study. Approximately half were enrolled on conventional computing tracks such as computer science, data science, immersive computing, robotics, computing for games, and game programming, among others ($N = 67$). The remaining half were on tracks adjacent to computing ($N = 66$) including digital art and animation, music technology, production, writing, virtual reality, and others. All these students were taught agile project management (see [23]) and Git version control in the module, as well as interpersonal conflict management and teamworking skills. Each student contributed to the development of their digital game in a practical way (i.e., contributing to the repository and build). Many respondents identified as male (72%), with 15% identifying as female, and 7.6% identifying as other, with the rest (5.3%) preferring not to state. The mean age was 20.7 years, with a majority (79%) within the 18–21 age bracket, and the remaining being older (22–45). This reflects the demographics within the department as a whole. The courses invited to take part in this survey have entry requirement of 112-120 Universities and College Admissions System (UCAS) points, and almost half of the participants have some prior experience of software development before joining the department.

4 RESEARCH METHODOLOGY

The research was conducted by means of a survey in the 2021-22 academic year. A single questionnaire was distributed to all students on the first-stage group project module. This was done at the end of the module, and remained open to students until an annual expo event in which they demonstrated the projects they had produced to the public. Students were invited to participate by email, were reminded of the study in live video updates by senior members of the department, and followed-up throughout the expo event.

The questionnaire[3] featured 46 six-point forced-choice Likert-type items. These were adapted from prior qualitative work [32] and extended through a four stage process: (i) interviews with four groups following *q*-methodology [31] to unpack the concepts previously identified, which is a method used to explore varying attitudes and evoke discussion by having participants rank and sort statements according to a grid; (ii) analysis and discussion by the authors to devise an initial set of questions aligned to the themes identified; (iii) three focus groups to refine the questions; and (iv) scrutiny from three experienced group supervisors to further refine the questions and establish face validity. The remaining questions collected data on demographic variables and complementary measures such as self-reported effort. Data was collected online using the SoScience[4] survey platform. This platform indicated that degradation in data quality due to the length of the questionnaire was minimal, according to a measure based on the relative speed index (see [26]). The questionnaire was completed with a median time of 22 minutes, with more than 60% of participants finishing within 30 minutes. There were 25 outliers, which either completed in less than 7 minutes or in more than 90 minutes. These responses were carefully scrutinised. No aberrations in response patterns were found. One incomplete response was excluded.

An exploratory factor analysis was conducted using SPSS following the maximum-likelihood method with a direct oblimin rotation and Kaiser normalisation. To aid in attaining a parsimonious set of items, some items were culled in a step-wise fashion according to the method described by Kano and Harada [22]. Extraction of factors was lead by the data, based on both the eigenfactor criterion. Further to the stepwise procedure, factors without strong loadings or with complex cross-loadings were removed. As a final step, where a factor had more than four items which had a clean and strong loading, only four items were kept in the model. These were selected based on appropriateness in discussion between the authors with the aim to maintain coverage of key aspects of the factor and to keep questions with the clearest phrasing.

5 RESULTS & ANALYSIS

5.1 Exploratory Factor Analysis

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (.711) indicated that the dataset was sufficient for factor analysis, but "middling" [20]. A review of the diagonal elements of the anti-image correlation matrix indicated this had skewed low due to just one item having a value below 0.5. Bartlett's Sphericity Test was also statistically significant ($\chi^2 = 805.391, df = 210, p < .001$). Thus indicating sufficient redundancy in the dataset for reduction. During the analysis, the authors converged on a seven-factor solution that utilised 21 of the items. The eigenvalue greater than one criterion indicated that a seven-factor model would be appropriate, and this was visually confirmed in the scree plot. A fit test showed that the reproduced covariance matrix was not statistically significantly different to the covariance matrix in the data ($\chi^2 = 93.612, df = 84, p = .222$) providing some evidence to support the goodness of fit of the seven-factor model.

The factors identified in the model include: (i) IMPACT, the level of endorsement of the beneficial impact of peer evaluation; (ii) ARBIT, to what extent arbitrary influences or bias were perceived to influence the peer evaluations; (iii) CONSIST, the

[1] https://moodle.org/
[2] https://feedbackfruits.com/group-member-evaluation
[3] http://repository.falmouth.ac.uk/4615/1/peer-evaluation-questionaire-2022.pdf
[4] https://www.soscisurvey.de/

Table 1: Pattern Matrix from a Maximum-Likelihood Exploratory Factor Analysis with Direct Oblimin Rotation Identifying Seven Key Factors Associated with Student Attitudes Towards Peer Evaluation

ID	Item	% Variance Explained	Cronbach's α	Factor Pattern Matrix						
				IMPACT	ARBIT	CONSIST	COHERE	REWARD	EASE	PROF
Impact		27.5	.835							
IMPACT-1	Peer evaluation helps lead us to a better project outcome			.822						
IMPACT-2	Peer evaluation helps my team to remain disciplined			.766						
IMPACT-3	Peer evaluation helps us to structure communication within the team			.701						
IMPACT-4	Peer evaluation improves the way we integrate into our teams			.615						
Arbitrary Influence		10.1	.759							
ARBIT-1	I have seen people use peer evaluation to negatively smear people they don't like				.866					
ARBIT-2	My peers use the evaluation process to complain about petty things				.676					
ARBIT-3	Factors beyond the project (e.g., popularity) influences the evaluations people receive from their peers				.609					
Consistency		8.5	.459							
CONSIST-1	Peer-evaluations follow the same schedule					.763				
CONSIST-2	Every team gets the same peer-evaluation experience					.345				
CONSIST-3	Supervisors do not run peer-evaluation in the same way*					-.339				
Team Coherence		7.5	.558							
COHERE-1	I work well with all members of my team						.648			
COHERE-2	Team meetings often turn into arguments*						-.528			
COHERE-3	My current team was able to get past the storming stage						.420			
Reward		6.5	.691							
REWARD-1	If I receive positive evaluations from my peers, that should mean I receive a high grade							.924		
REWARD-2	My peers should never be able to influence the marks I receive*							-.442		
REWARD-3	It is sensible to link the peer evaluation and assessment processes							.301		
Ease		5.2	.563							
EASE-1	Peer-evaluation is an unreasonably slow and tiring process*								-1.009	
EASE-2	It does not take me too long to do peer-evaluations								.409	
EASE-3	I get my peer-evaluations done on-time								.292	
Professionalism		4.6	.454							
PROF-1	Peer evaluation enables me to assert my real views									.675
PROF-2	I feel compelled to keep evaluations of my peers positive to keep the team working smoothly*									-.502

* denotes an item with a reversed valence

perceived consistency of the process; (iv) COHERE, to what extent groups became able to function in a coherent manner; (v) REWARD, sentiment towards the influence of peer evaluation on the assessment process and whether it should be linked to grading; (vi) EASE, the ease of the process; and (vii) PROF, the degree with which the team engaged in an open and honest dialogue and avoided shaping their engagement in unprofessional ways (e.g., not giving their real views in an attempt to keep people happy).

Table 1 shows items included in the analysis following the final iteration of data reduction, illustrating how each question maps to its underlying factor. It also reports their relative proportion of explained variance and Cronbach's alpha internal consistency measure of reliability for each factor. Cross-loadings below 0.2 have been suppressed. This model explained 50.3% of the variance. It is worth noting that many of themes hypothesised when designing the questionnaire were unidimensional, with 21 items affiliated with the factor now labelled IMPACT, reduced to just four in the final model. This left most of the factors with at least three items with strong loadings. However, notable exceptions were CONSIST, REWARD, and EASE which each had one item with a weaker loading. Furthermore, the EASE factor had one item with a weak loading and also a cross-loading with ARBIT, whilst the PROF factor has two items with weak loads, one of which cross-loaded with IMPACT.

5.2 Correlation Analysis

A Spearman's correlation analysis is shown in Table 2. It shows several relationships between the factors are statistically significant. Those students tending to endorse the coherence of their team, establishing a link between peer evaluation and assessment, as well as the ease of the process also tended endorsed the impact of peer evaluation. Notably, it seems that those observing high levels of arbitrary influence tended not to endorse the impact of peer evaluation. They also tended to not to endorse the consistency of the process or establish a coherent team. Arbitrary influence is also negatively correlated with attainment. There seems to be a link between the ease of the process and its consistency. Those endorsing they were in a coherent team were more likely to endorse linkages between peer evaluation and assessment. They also seemed to observe higher levels of professionalism. Team coherence was also positively correlated with attainment. There seems to be a link between attitudes towards rewards and the ease of the process.

Table 2: Spearman's Correlation Analysis Showing the Relationship with Self-Reported Effort and Attainment

	IMPACT	ARBIT	CONSIST	COHERE	REWARD	EASE	PROF	Effort	Attainment
IMPACT	–								
ARBIT	-.259**	–							
CONSIST	-.038	-.244**	–						
COHERE	.312**	-.371**	.028	–					
REWARD	.355**	-.088	.125	.187*	–				
EASE	.279**	-.149	.211*	.078	.214*	–			
PROF	.067	-.032	.086	.235**	.031	.247**	–		
Effort	-.065	.010	.015	.119	.001	.017	-.008	–	
Attainment	.118	–.162*	-.077	.293**	.181*	-.109	.184*	.382**	–

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Additionally, those students endorsing this link also seemed to attain higher marks. Those experiencing ease with the process seemed to observe higher levels of professionalism in their team. Finally, the professionalism of the team and the effort an individual dedicated to their project were both positively correlated with attainment.

6 DISCUSSION

6.1 From the student perspective, in which key ways does their experience of peer evaluation vary?

There are many perceived benefits to peer evaluation, so much so that eighteen of the original items corresponded to this factor. Beyond the four in the final model, these included: helping the team remain disciplined; improving productivity; helping the team towards a better outcome; helping individuals contribute better to the team; improving team dynamics and team member integration; critiquing development practices; setting expectations and manage a realistic workload; nurturing professionalism; highlighting concerns; structuring and otherwise improving communication; pushing the team to develop skills it is missing; increasing engagement; focusing attention on key milestones and processes, alongside others.

These complement findings in prior work on the perceived purpose of peer evaluation [32]. However, the key contribution of this work is unveiling those factors along which the experience of peer evaluation varies. There were two strong themes which emerged in the qualitative data and were further confirmed through this quantitative data. These were the potential for arbitrary influence to diminish the process and concerns about consistency. The level of concern varies across students, but is nonetheless a facet of peer evaluation that educators will need to address to reassure students of the efficacy of the process under a range of possible vulnerabilities.

Team coherence is another theme that emerged corresponding to the earlier qualitative data. These included a teams ability to function well enough to be able to exploit the benefits of the peer evaluation. This decomposed into those teams that were able to establish norms and start performing in a timely manner, potentially reflecting whether or not a team was able to reach a state where they were able to adapt to feedback effectively. The alternative, perhaps, being teams that stuck in the forming or storming stages of Tuckman's team development sequence [47]. That is, respectively, where members of the team do not know each other or work closely enough to engage in peer evaluation, and where members of the team experience friction and engage in conflict undermining their ability to evaluate the team and act on observed weaknesses.

Related to this previous theme, is the professionalism exhibited within a team. When teams are not honest with their appraisals this would undermine the peer evaluation. It seems that there is variance on when the students feel able to assert their real views, instead feeling compelled to distort their feedback to keep the team working smoothly, despite the fact that such feelings could undermine the process and be counter-productive. Thus, this aspect is important for educators to consider. The final factors which influenced the way in which peer evaluation was experienced was attitudes towards the way peer evaluation could influence the grades being awarded. Students varied in their views considerably regarding this. It will be important to be transparent with students to ensure there is adequate explanation of how the process intersects with the ways in which grades are arrived at.

6.2 To what extent do these factors correlate with effort, attainment, and with each other?

Two key findings are highlighted in the correlation analysis. Firstly, there is not a correlation between attainment and perceived impact of peer evaluation. This lack of correlation indicates that students who believed peer evaluation activities had value were not necessarily using the process to improve their performance and vice-versa, those not valuing it were still making use of it. This may indicate that the way peer evaluation is being conducted is not translating into actionable feedback for student improvement. It could also indicate that students are primarily drawing upon other factors outside of the peer evaluation process to improve their own practice, rather than using the process as intended. This could indicate an unobserved variable related to the political side of team dynamics causing the peer evaluation process not to be applied as intended by the tutors, and point towards possibilities future studies should examine. In addition, this is somewhat

concerning as one of the key reasons for conducting and assessing the peer evaluation process is to help students genuinely improve their practice. The questions relating to impact only deal with student experience of the peer evaluation process, it is possible that students are using this process to improve, but this is not being perceived by the students.

A second key finding from the correlation analysis is that there is not a correlation between the professionalism factor and the consistency factor, nor between consistency and attainment. Prior work [32] reports that students care deeply about parity in terms of process. The statements about consistency were primarily concerned with the belief that team members got the same treatment both within their teams and across different supervisors. This could indicate that students do not believe that there is a strong tie between team members' engaging honestly with the peer evaluation process and the process ensuring that everyone gets a 'fair shout'. It is interesting that students, often primarily concerned with consistency of an academic experience, themselves approach these processes in an inconsistent manner. This observation highlights the importance of renewing a focus on the honesty and transparency of team discussions, especially in the presence of tutors who are seen as guardians of the process. Another possible interpretation is that the wording of the questions meant that students made use of second hand accounts from others when considering the question of consistency, with students reflecting on their own experience in the professionalism questions, but more broadly (based on reports from their colleagues in other groups) with the consistency questions.

6.3 Implications for Practice

Those educators developing and deploying peer evaluation methods cannot ignore or gloss over the socio-cultural contexts in which peer evaluation methods operate. The most successful students claimed to have a coherent team capable of professional dialogue with minimal arbitrary influence. This suggests that efforts to help students adopt a professional approach, to minimise arbitrary influence, and to actively help teams to progress past the storming phase of Tuckman's model [47] in a timely manner attain greater teamworking skills than those who do not. In addition, a peer evaluation process could have a role in helping the team to converge on productive approaches to collaboration. Though consistency did not relate to attainment, it is nevertheless of considerable concern to students. Therefore, it is important that educators address any inconsistencies in the peer evaluation process to ensure parity and a clear understanding of the benefits.

From the analysis, the desire for consistency is not wholly driven from the use of peer evaluation as a marking tool. This would hint that it comes from some underling desire for fairness in the peer evaluation process. As a result, it is important to ensure that the process is seen as fairly and consistently implemented, even if peer evaluation is being used for formative purposes.

Most of the factors identified in this paper correlate with the impact factor. This implies that the student perceptions of the impact of peer review can be improved by ensuring the process is consistent across teams, free of arbitrary influence and ensuring the process is not too tiring. One possible way to address this is to ensure that students are aware of the process of providing suitable feedback, to help minimise arbitrary influences. Another way this can be improved is to reduce the burden of running peer evaluations on students by ensuring the process is a streamlined as possible, such as not having too many criteria and ensuring the process doesn't require too much time to complete. For large teams one approach for reducing this burden is to only require each team member to evaluate a subset of the other members. As the impact factor is also correlated with reward, students may perceive the peer evaluation as having more impact if it influences the marks.

7 CONCLUSION

This paper establishes a conceptual framework for investigating peer evaluation. Many questions around the design of this framework remain, but it is clear it should address factors including: the perception of any beneficial impacts; arbitrary influences; procedural consistency; the ability of a team to function coherently and thereby act upon feedback from evaluations; the link between peer-evaluation and assessment processes; the ease of the process; and to what extent teams can engage in open and honest dialogue. Most of these constructs demonstrate good characteristics in terms of reliability and validity. However, more work is needed to develop a robust measurement scale. In particular, the professionalism factor seems important, but is deserving of more attention to improve measurement quality. Such enhancements, would pave the way for comparing the efficacy of peer evaluation methods.

The findings suggest links between peer evaluation and attainment which go beyond effort. Students care passionately about the critique they receive and acknowledge the many positive impacts it can lead to. However, it is important that educators streamline peer evaluation and help their students to establish coherent teams capable of professional discourse. To this end, considering means of minimising arbitrary influence. It is also important to consider how to facilitate the workload introduced by the need to produce highly constructive feedback and establish a consistent approach across teams to ensure fairness.

It is important to acknowledge the exploratory nature of this study. Though it is building upon prior qualitative work using new quantitative data, further replication is warranted. Replication will indicate the generality and validity of the model in different contexts, as this work was conducted at a single institution. There could be confounding factors specific to the cohort studied, such as acquiescence bias because some of the authors were involved in the particular module studied. Further investigation is also needed to explore the applicability of the framework across demographics, contexts, and project types, making further considerations for equality, diversity, and inclusion.

REFERENCES

[1] Efthimia Aivaloglou and Anna van der Meulen. 2021. An Empirical Study of Students' Perceptions on the Setup and Grading of Group Programming Assignments. *ACM Trans. Comput. Educ.* 21, 3, Article 17 (mar 2021), 22 pages. https://doi.org/10.1145/3440994

[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. 1995. Assessment and learning: contradictory or complementary. *Assessment for learning in higher education* (1995), 35–48.

[4] 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

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

[6] 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.

[7] 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.

[8] 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.

[9] Nicole Clark, Pamela Davies, and Rebecca Skeers. 2005. Self and peer assessment in software engineering projects. In *Proceedings of the 7th Australasian conference on Computing education-Volume 42*. 91–100.

[10] Tom Crick, Tom Prickett, Jill Bradnum, and Alan Godfrey. 2022. Gender parity in peer assessment of team software development projects. In *Computing Education Practice 2022*. 9–12.

[11] Jessica A Crowe, Tony Silva, and Ryan Ceresola. 2015. The effect of peer review on student learning outcomes in a research methods course. *Teaching Sociology* 43, 3 (2015), 201–213.

[12] 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

[13] Nancy Falchikov. 1995. Peer feedback marking: Developing peer assessment. *Innovations in Education and training International* 32, 2 (1995), 175–187.

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

[15] 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.

[16] Johan Gärdebo and Mattias Wiggberg. 2012. *Students, the university s unspent resource: Revolutionising higher education through active student participation.*

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

[18] 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/

[19] Nicole Herbert. 2007. Quantitative Peer Assessment: Can Students Be Objective?. In *Proceedings of the Ninth Australasian Conference on Computing Education - Volume 66* (Ballarat, Victoria, Australia) *(ACE '07)*. Australian Computer Society, Inc., AUS, 63–71.

[20] Henry F Kaiser. 1974. An index of factorial simplicity. *psychometrika* 39, 1 (1974), 31–36.

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

[22] Yutaka Kano and Akira Harada. 2000. Stepwise variable selection in factor analysis. *Psychometrika* 65, 1 (2000), 7–22.

[23] Clinton Keith. 2010. *Agile game development with Scrum.* Pearson Education.

[24] Ilana Lavy and Aharon Yadin. 2010. Team-Based Peer Review as a Form of Formative Assessment - The Case of a Systems Analysis and Design Workshop. *Journal of Information Systems Education* 21, 1 (2010), 85.

[25] 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.

[26] Dominik Johannes Leiner. 2019. Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. In *Survey Research Methods*, Vol. 13. 229–248.

[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] Robert Lingard and Shan Barkataki. 2011. Teaching teamwork in engineering and computer science. In *2011 Frontiers in education conference (FIE)*. IEEE, F1C–1.

[29] 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

[30] Andrew Luxton-Reilly. 2009. A systematic review of tools that support peer assessment. *Computer Science Education* 19, 4 (2009), 209–232.

[31] Bruce McKeown and Dan B Thomas. 2013. *Q methodology*. Vol. 66. Sage publications.

[32] Alexander Mitchell, Terry Greer, Warwick New, Joseph Walton-Rivers, Matt Watkins, Douglas Brown, and Michael James Scott. 2021. Student Perspectives on the Purpose of Peer Evaluation During Group Game Development Projects. In *United Kingdom and Ireland Computing Education Research Conference*. (Glasgow, United Kingdom) *(UKICER '21)*. Association for Computing Machinery, New York, NY, USA, Article 7, 7 pages. https://doi.org/10.1145/3481282.3481294

[33] Catherine Moore and Susan Teather. 2013. Engaging students in peer review: Feedback as learning. (2013), 16.

[34] 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

[35] Raoul A Mulder, Jon M Pearce, and Chi Baik. 2014. Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education* 15, 2 (2014), 157–171.

[36] 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

[37] 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

[38] Quality Assurance Agency. 2014. The Frameworks for Higher Education Qualifications of UK Degree-Awarding Bodies. https://www.qaa.ac.uk/docs/qaa/quality-code/qualifications-frameworks.pdf

[39] P Race. 1998. Practical pointers on peer assessment. *SEDA PAPER* (1998), 113–122.

[40] 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

[41] 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.

[42] Jacklin Stonewall, Michael Dorneich, Cassandra Dorius, and Jane Rongerude. 2018. A review of bias in peer assessment. In *2018 CoNECD-The Collaborative Network for Engineering and Computing Diversity Conference*.

[43] Thyago Tenório, Ig Ibert Bittencourt, Seiji Isotani, and Alan Pedro Silva. 2016. Does peer assessment in on-line learning environments work? A systematic review of the literature. *Computers in Human Behavior* 64 (2016), 94–107.

[44] Robert Thompson. 2001. Reliability, Validity, And Bias In Peer Evaluations Of Self Directed Interdependent Work Teams. In *2001 Annual Conference*. 6–845.

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

[46] Yanbin Tu and Min Lu. 2005. Peer-and-Self Assessment to Reveal the Ranking of Each Individual's Contribution to a Group Project. (2005), 10.

[47] Bruce Tuckman. 1965. Developmental sequence in small groups. *Psychological Bulletin* 63, 6 (1965), 384–399.

[48] Marjo Van Zundert, Dominique Sluijsmans, and Jeroen Van Merriënboer. 2010. Effective peer assessment processes: Research findings and future directions. *Learning and instruction* 20, 4 (2010), 270–279.

[49] 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

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

[51] 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 (2010), 429–443.

[52] James Wood. 2021. Making peer feedback work: the contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment & Evaluation in Higher Education* (2021), 1–20.

[53] Qiyun Zhu and David Carless. 2018. Dialogue within peer feedback processes: clarification and negotiation of meaning. *Higher Education Research & Development* 37 (March 2018), 1–15. https://doi.org/10.1080/07294360.2018.1446417