Gender parity in peer assessment of team software development projects

Anonymous Author(s)

ABSTRACT

Development projects in which small teams of learners develop software/digital artefacts are common features of computing-related degree programmes. Within these team projects, it can be problematic ensuring students are fairly recognised and rewarded for the contribution they make to the collective team effort and outputs. Peer assessment is a commonly used approach to promote fairness and due recognition. Maintaining parity within assessment processes is also a critical aspect of fairness. This paper presents the processes employed for the operation of one such team project at a UK higher education institution, using the Team-Q rubric and analysing the impact of the (self-identified) gender of learner marking and the learner being marked on the scores obtained. The results from this institutional sample (N=121) using the Team-Q metric offers evidence of gender parity in this context. This study also makes the case for continued vigilance to ensure Team-Q and other rubrics are used in a manner that supports gender parity in computing.

CCS CONCEPTS

• **Social and professional topics → Computing education**; **Student assessment**.

KEYWORDS

Peer assessment, group projects, team working, gender, diversity

ACM Reference Format:
Anonymous Author(s). 2022. Gender parity in peer assessment of team software development projects. In *CEP '22: ACM Computing Education Practice Conference, January 06, 2022, Durham, UK*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

1 WHAT IS IT?

In the UK and other jurisdictions, computing degrees regularly contain courses in which teams of learners develop software/digital artefacts. This commonly involves the demonstration of various software engineering competencies, for example; analysis and design; implementation; testing; configuration management and version control; team working; project management/control, etc. Such teamwork projects are not universally well-received or well-regarded by students [9]; however there is an employability-related dimension [12, 22] to the development of these competencies and they are required characteristics by professional bodies for degree accreditation [5].

Within these team projects, all members of the teams are expected to contribute to the development of these software/digital artefacts. Sometimes the contribution is structured in some manner by task, by role, etc; in other projects, the teams are more self-managing. Commonly, there will be some expected collective outcomes. These outcomes could be a product (e.g. common look and feel, an integrated product, etc) or task-based (e.g. risk analysis, project plan, team demonstration of the product, etc). In addition to the collective tasks there may be individual tasks which again could be product (e.g. building of sub-system X) or task-based (e.g. testing of the product). Assuring fair contributions from all learners to collective tasks can present challenges [15]. One common approach to address this challenge is peer assessment [9]. It has been reported that learners appreciate that peer assessment provides a mechanism to hold members accountable and aid fairer marking. It also supports giving and receiving feedback, prompts personal reflection and improvement, supports supervision, informing project planning and management, facilitates exploring and reshaping group dynamics, improves project outputs, and offering a sense of safety to raise issues without repercussions [12]. Peer assessment schemes which are based upon peer ranking and peer rating systems [23] have been proposed. One common form of peer rating system is for learners to assess their peers by a given metric, calculating means for individual and teams, and then weighting collective marks accordingly. This can be achieved by online tools such as WebPA [14], BuddyCheck.io [18], or SparkPlus [19] or via the use of online surveying tools. In general terms, the algorithm commonly adopted is then: (*i*) each learner scores each of their peers in their team by a set of metrics; (*2*) a weighting is then calculated by Mean Peer Score for the learner divided by Mean Peer Score for the whole team; and (*3*) the individual learner would then be awarded the mark that team is awarded multiplied by this weighting. Such team projects commonly have individual assessed outcomes as well as team responsible assessed outcomes and it would normally only be the team outcomes that would be subject to the weighting.

Clearly the criteria used has a significant influence on the weighting and hence potentially the grade awarded to the individual learners. Given the number of individuals who contribute to the marking (i.e. all the learners) there is more potential for unconscious or conscious bias to influence the marking than if the marking was completed solely by faculty. This paper explores whether a validated approach – Team Q [2] – together with a specific set of practices, supports gender parity in terms of the peer assessment marks awarded.

Team Q [2] assesses five components of team working: contributes to team project; facilitates contributions of others; planning and management; fosters a team climate; and manages potential

CEP '22, January 06, 2022, Durham, UK
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

conflict. Each of the components is in turn measures by indication of '*how often does your peer demonstrate the following*' against a set of descriptions. Each description is awarded scores as follows: Never=0; Sometimes=1; Usually=2; Regularly=3; and Always=4. The full set of descriptions can be seen in Table 1. Hence a learner scores each of their peers out of 56 overall. The peer weighting is then calculated using the algorithm indicated previously (with a learner's mean peer score divided by their team's mean peer score).

The Team Q peer model produces a team work weighting via a marking scheme synthesised from wider research. In so doing, it presents a comprehensive model for what constitutes effective team working. One of the outcomes is to highlight to learners a rounded model of what competencies constitute good quality team working and these are far from solely technical competencies. The authors contend this provision of a benchmark for good team working practice provides useful formative feedback for the learners as they complete the projects in their teams.

2 WHY ARE YOU DOING IT?

It is widely recognised that Science, Technology, Engineering and Mathematics (STEM) (e.g. [1]) and more specifically computing remain male-dominated disciplines. In the UK, only 1 in 5 Computing, Engineering and Technology students were female in the 2019-2020 academic year [11]. For the computing discipline in the UK, 26,285 out of 105,485 (just less than 20%) identified as female and 210 identified as non-binary [11]. Addressing this imbalance is critical for the disciplines involved, from a social, cultural and economic perspective, to maximise the potential future development of the discipline. This is further reinforced by the ambitions of the United Nations Sustainable Development Goals: *SDG4: Quality Education* and *SDG5: Gender Equality*.

Belonging [25] is recognised as a crucial factor for retaining learners within the computing discipline, yet learners who self-identify as women have been reported to have a lower sense of belonging [13]. The challenges faced by female students related to belonging has also been explored by qualitative studies [26]. Together, this highlights the need to carefully evaluate whether education practices promote belonging, support diversity, and gender parity. Whilst there is always the need to assure equity in assessment, in this case when learners are contributing to the assessment processes of their peers the need to assure the processes used exhibits gender parity is particularly strong.

Peer assessment is the main focus of this work; however, it is acknowledged the data is being gathered as part of assessed tasks which may, in some way, influence the outcomes. Additionally, more details of the full practices adopted are provided as different results may arise if the scheme was applied in alternative assessment situations.

3 WHERE DOES IT FIT?

The team project is run as part of the final year of an undergraduate Computer Science degree at a UK higher education institution. The study took place in the 2020-2021 academic year, and thus under the constraints of the COVID-19 pandemic. The course runs between January and April over a 15-week period including a 3-week spring vacation. Institutional ethical approval for the study was obtained. Explicit written consent was obtained as part of the peer assessment learners were asked to indicate "What gender do you identify as?" as an optional free text field; additionally, they were specifically requested to approve their consent to be included in the study. Hence learners have the opportunity to not supply the information and additionally specifically consent to participate. Those who provided a null response have not been included in the study. The size of the cohort was 170. Of this group, 121 learners are included in the study with 108 learners self-identifying as male ('male' or 'man' or 'masculine') and 13 self-identifying as female. A further three learners identified responded as non-binary (responding: 'I don't know'; 'non-binary'; and 'nothing'); these have not been included in the study due to concerns that they may be individually identifiable.

The key aspects of the management of the projects are as follows:

Team allocation: The learners are allowed to either self-select their teams or choose to be assigned a team. Teams are normally comprised of five individuals. If learners wish to be assigned a team, these are allocated randomly upon a first-come, first-served basis.

Project selection: All the projects are 'live' development projects in the sense that teams develop a software/digital artefact for a third-party. Some of the projects are self-sourced by the learners. Additionally, the tutors assist some teams in identifying suitable projects.

Learning Agreements: As part of the establishment of teams, the learners are required to produce a learning agreement which documents key decisions regarding how the team will collaborate to complete the work. Teams are encouraged to reflect upon this agreement as the projects progress. A writing frame is provided posing key questions the learning agreement should address.

Development approach: The teams are required to follow a full-stack development approach with each team member developing a subsystem that they can ultimately demonstrate individually if required. However, the teams are encouraged to demonstrate a fully-integrated working product and are recognised for doing so as part of the marking rubric. If presenting an integrated working product is not possible for reasons beyond an individual learner's control (for example, there is a passenger in the team) adjustments are made so that a learner is not unfairly penalised.

Support: The teams are supported by weekly progress review meetings with a tutor. These follow a stand-up style with each team member asked to identify progress and any road blocks which can then be discussed in more depth. A Microsoft Word and a InVision Freehand template were provided to support this activity. These records were uploaded to the institutional virtual learning environment at the end of meetings. External to the meetings, the supervising tutor attempted to support the teams to resolve any team-related issues. For a small number of groups this involved removing a team member for serial non-engagement with either the supervision meetings or, more importantly, lack of engagement with the team.

Assessment: There are three related components of summative assessment: a project proposal (10%), a demonstration of the software (50%) and a report which critically evaluates the project and the professional, ethical, legal and social issues a finalised and deployed version of the produced prototype would need to mitigate. The team aspects are: 50% of the proposal and 20% of the

Table 1: Means of Team Q Score by gender of marked learner and by gender of marker pairing (female marking female, female marking male, male marking female and male marking male)

		Marked Gender		Marker Gender / Marked Gender			
Component	Description	Female	Male	Female-Female	Female-Male	Male-Female	Male-Male
	Mean Team-Q Score	46.94	47.15	50.60	46.41	46.30	47.22
	Number of marks awarded in each category	32	323	5	27	27	296
Contribute to team Project	*Participates actively and accepts a fair share of the group work*	3.47	3.49	3.80	3.44	3.41	3.49
	Works skilfully on assigned tasks and completes them on time	3.44	3.37	3.40	3.19	3.44	3.38
	Gives timely, constructive feedback to team members, in the appropriate format	3.19	3.24	3.20	3.26	3.19	3.23
Facilitates contributions of others	*Communicates actively and constructively*	3.28	3.38	3.60	3.37	3.22	3.38
	Encourages all perspectives be considered and acknowledges contributions to others	3.41	3.44	3.80	3.37	3.33	3.44
	Constructively builds on the contributions of others and integrates own work with work of others	3.38	3.36	3.80	3.33	3.30	3.37
Planning and Management	*Takes on an appropriate role in the group (e.g. leader, note take, etc)*	3.31	3.05	4.00	2.93	3.19	3.06
	Clarifies goals and plans the project	3.22	3.20	3.60	3.00	3.15	3.22
	Reports to team on progress	3.38	3.35	3.60	3.30	3.33	3.34
Fosters a team climate	*Ensures consistency between words, tone, facial expressions, and body language*	3.47	3.47	3.60	3.63	3.44	3.46
	Expresses positivity and optimism about team members and project	3.44	3.49	3.40	3.63	3.33	3.48
Manages potential conflict	*Displays appropriate assertiveness: neither dominating, submissive nor passive aggressive*	3.34	3.42	3.60	3.30	3.30	3.43
	Contributes to appropriately healthy debate	3.28	3.45	3.60	3.30	3.22	3.46
	Responds to and manages direct/indirect conflict constructively and effectively	3.44	3.46	3.60	3.37	3.41	3.46

demonstration which are marked as a team and are weighted by peer assessment.

Peer Assessment: The project proposal and the demonstration contain team and individual tasks. As such peer assessment is employed to ensure a fair split of marks between the team. Over various historical deliveries of the course various technologies have been used to administer the peer assessment including paper, virtual learning environment tools and other electronic tools. In this delivery, peer assessment was administered by Microsoft Forms. Two rounds of peer assessment were completed, one formative and one summative. Only the summative round is included in the study.

4 DOES IT WORK?

The responses to the Team-Q rubric were analysed using a combination of Excel and R (v4.1.0). Excel was primarily used for data storage and cleaning; R was used for the statistical analysis.

The Team-Q Score, number of marks awarded, and means for responses to the different descriptions in the Team-Q rubric by the gender of the marked learner and by the gender of the marker pairing are shown in Table 1 above. A t-test is indicative of there being no evidence of statistical difference in the mean of Team-Q Score for the marks awarded to female (46.94) and male (47.15) leaners (t=-0.087708, df=35.438, p=0.9306). Analysis of variance (ANOVA) suggests little statistical difference in the mean Team-Q Score awarded between "female-marking male" (46.41), "male-marking female" (46.30) or "male-marking male" (47.22) pairs (markers gender F=0.104, p=0.748 and marked gender F= 0.177, p=0.674). The slightly higher female-to-female marking pairing mean (50.60) is not statistically significantly different to the other pairings as can be seen from a t-test (t=0.697, p=0.487). Together this provides confidence that Team-Q exhibits gender parity in terms of the gender of the learner being marked and gender of the learner completing the marking. This is a sample size of N=121 learners on one course delivered with a low incidence of female learners (13) but even so the results are encouraging.

5 WHO ELSE HAS DONE THIS?

Peer assessment and related web-based peer assessment has been advocated as a mechanism for equitable assessment of contribution to team and team software development projects for a number of years [2, 4, 7, 9, 15, 16]. It has also been reported that, when it is used in a summative context, there can be bias due to affiliation with a group [3], and learners sometimes do not want to award a low mark to their peers (and, understandably, particularly to their friends) [20]. Bias in peer assessment on the basis of gender has been widely reported [10, 21], and elsewhere bias has not been evidenced [8, 24]. This mixed picture highlights the need to validate tools employed in different contexts to assure the process exhibits gender parity.

6 WHAT WILL YOU DO NEXT?

It is possible that the rapid shift to online learning, teaching and assessment during the COVID-19 pandemic [6] may have influenced the results, and as such the intention is to repeat the study with this year's cohort to determine if the results are reproducible in more typical learning conditions. There is the potential to extend the study to other courses at the university, as well as other institutions to explore whether the outcomes are reproducible in different circumstances. The focus of the study to date has overlooked non-binary learners due the risks related to identification of individual learners. Consideration needs to be made for how the impact upon non-binary learners can be explored. It has been reported that self-identified minorities can have a lower sense of belonging [13]. Learners may identify as minorities for reasons other than gender (e.g. ethnicity, neurodiversity, no family history with higher education, etc) and considering such factors is an area for future work. Finally, since there are performance benefits associated with diverse teams [17], exploring the impact of team diversity upon peer assessment and overall achieved grades is an area for further exploration.

7 WHY ARE YOU TELLING US THIS?

It is encouraging that there was evidence of gender parity within the peer assessment scheme adopted for this study. However, this is for one cohort at one university, using a particular set of processes for work completed during the COVID-19 pandemic. As such, a wider investigation into different contexts and whether the Team-Q rubric continues to be equitable. Arguably, Team-Q is a well-established rubric which explores more dimensions of team working than some of the more standard approaches that are embedded in existing tools [14, 18, 19]. Although such tools could easily be configured to use the Team-Q rubric or other scheme as an alternative to their default. Furthermore, given the low overhead of evaluating the impact of self-identified gender upon peer assessment results, doing so is a practical recommendation for the occasions when peer assessment is employed.

Finally, gender parity is not the only possible dimension of parity that should be exhibited in peer assessment and other assessment approaches. This points to a rather urgent set of work to ensure assessment approaches are equitable for different demographics (e.g. ethnicity, neurodiversity, etc).

REFERENCES

[1] Chardie L. Baird. 2018. Male-dominated stem disciplines: How do we make them more attractive to women? *IEEE Instrumentation Measurement Magazine* 21, 3 (2018), 4–14. https://doi.org/10.1109/MIM.2018.8360911

[2] Emily Britton, Natalie Simper, Andrew Leger, and Jenn Stephenson. 2017. Assessing teamwork in undergraduate education: a measurement tool to evaluate individual teamwork skills. *Assessment & Evaluation in Higher Education* 42, 3 (2017), 378–397. https://doi.org/10.1080/02602938.2015.1116497

[3] 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

[4] Nicole Clark, Pamela Davies, and Rebecca Skeers. 2005. Self and Peer Assessment in Software Engineering Projects. In *Proc. 7th Australasian Conf. on Computing Education.* 91–100.

[5] Tom Crick, James H. Davenport, Paul Hanna, Alastair Irons, and Tom Prickett. 2020. Computer Science Degree Accreditation in the UK: A Post-Shadbolt Review Update. In *Proc. of CEP'20.* ACM, Article 6. https://doi.org/10.1145/3372356.3372362

[6] Tom Crick, Cathryn Knight, Richard Watermeyer, and Janet Goodall. 2020. The Impact of COVID-19 and "Emergency Remote Teaching" on the UK Computer Science Education Community. In *Proc. of UKICER'20.* ACM. https://doi.org/10.1145/3416465.3416472

[7] Fabian Fagerholm and Arto Vihavainen. 2013. Peer assessment in experiential learning Assessing tacit and explicit skills in agile software engineering capstone projects. In *Proc. of FIE'13.* 1723–1729. https://doi.org/10.1109/FIE.2013.6685132

[8] Nancy Falchikov and Douglas Magin. 1997. Detecting Gender Bias in Peer Marking of Students' Group Process Work. *Assessment & Evaluation in Higher Education* 22, 4 (1997), 385–396. https://doi.org/10.1080/0260293970220403

[9] Neil Andrew Gordon. 2010. Group working and peer assessment — using WebPA to encourage student engagement and participation. *Innovation in Teaching and Learning in Information and Computer Sciences* 9, 1 (2010), 20–31. https://doi.org/10.11120/ital.2010.09010020

[10] Laura Heels and Marie Devlin. 2019. Investigating the Role Choice of Female Students in a Software Engineering Team Project. In *Proc. of CEP'19.* ACM, Article 2. https://doi.org/10.1145/3294016.3294028

[11] HESA. 2021. What do HE students study?: Personal characteristics. https://www.hesa.ac.uk/data-and-analysis/students/what-study/characteristics

[12] 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. ACM, Article 7. https://doi.org/10.1145/3481282.3481294

[13] Catherine Mooney and Brett A. Becker. 2020. Sense of Belonging: The Intersectionality of Self-Identified Minority Status and Gender in Undergraduate Computer Science Students. In *Proc. of UKICER'20.* ACM, 24–30. https://doi.org/10.1145/3416465.3416476

[14] Engineering Centre of Excellence in Teaching and Learning. 2005. WebPA. https://webpa.lboro.ac.uk/login.php

[15] Helen Phillips, Wendy Ivins, Tom Prickett, Julie Walters, and Rebecca Strachan. 2021. Using Contributing Student Pedagogy to Enhance Support for Teamworking in Computer Science Projects. In *Proc. of CEP'21.* ACM, 29–32. https://doi.org/10.1145/3437914.3437976

[16] Richard Raban and Andrew Litchfield. 2007. Supporting peer assessment of individual contributions in groupwork. *Australasian Journal of Educational Technology* 23, 1 (Mar. 2007). https://doi.org/10.14742/ajet.1272

[17] David Rock and Heidi Grant. 2016. Why are diverse teams smarter. *Harvard Business Review* (2016). https://hbr.org/2016/11/why-diverse-teams-are-smarter

[18] Shareworks. 2021. Buddy Check. https://www.buddycheck.io/

[19] SparkPlus. 2021. Introduction to SparkPlus. https://sparkplus.com.au/

[20] Baharini Sridharan, Joanna Tai, and David Boud. 2019. Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education* 77 (2019), 853–870. https://doi.org/10.1007/s10734-018-0305-7

[21] Jacklin Stonewall, Michael Dorneich, Cassandra Dorius, and Jane Rongerude. 2018. A Review of Bias in Peer Assessment. In *Proc. of CoNECD 2018.*

[22] Sarah. Thomas and Susan Busby. 2003. Do industry collaborative projects enhance students' learning. *Education + Training* 45, 4 (2003), 226–235. https://doi.org/10.1108/00400910310478157

[23] Yanbin Tu and Minn Lu. 2005. Peer-and-Self Assessment to Reveal the Ranking of Each Individual's Contribution to a Group Project. *Journal of Information Systems Education* 16, 2 (2005), 197–206.

[24] Richard Tucker. 2014. Sex does not matter: gender bias and gender differences in peer assessments of contributions to group work. *Assessment & Evaluation in Higher Education* 39, 3 (2014), 293–309. https://doi.org/10.1080/02602938.2013.830282

[25] Nanette Veilleux, Rebecca Bates, Cheryl Allendoerfer, Diane Jones, Joyous Crawford, and Tamara Floyd Smith. 2013. The Relationship between Belonging and Ability in Computer Science. In *Proceeding of SIGCSE'13.* ACM, 65–70. https://doi.org/10.1145/2445196.2445220

[26] Emily Winter, Lisa Thomas, and Lynne Blair. 2021. 'It's a Bit Weird, but It's OK'? How Female Computer Science Students Navigate Being a Minority. In *Proc. of ITiCSE'21.* ACM, 436–442.