Journal of Information Technology Education: Innovations in Practice

An Official Publication of the Informing Science Institute InformingScience.org

JITEiip.org

Volume 21, 2022

MATCHING AUTHORS AND REVIEWERS IN PEER ASSESSMENT BASED ON AUTHORS' PROFILES

Amal Alkhalifa*	Department of Computer Science and Information Technology, Applied College, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia	akalkalifh@pnu.edu.sa
Marie Devlin	School of Computing, Newcastle University, Newcastle upon Tyne, UK	marie.devlin@newcastle.ac.uk
Mona Alkhattabi	College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia	mkhattabi@imamu.edu.sa

* Corresponding author

ABSTRACT

Aim/Purpose	To encourage students' engagement in peer assessments and provide students with better-quality feedback, this paper describes a technique for author-reviewer matching in peer assessment systems – a Balanced Allocation algorithm.
Background	Peer assessment concerns evaluating the work of colleagues and providing feedback on their work. This process is widely applied as a learning method to involve students in the progress of their learning. However, as students have different ability levels, the efficacy of the peer feedback differs from case to case. Thus, peer assessment may not provide satisfactory results for students. In order to mitigate this issue, this paper explains and evaluates an algorithm that matches the author to a set of reviewers. The technique matches authors and reviewers based on how difficult the authors perceived the assignment to be, and the algorithm then matches the selected author to a group of reviewers who may meet the author's needs in regard to the selected assignment.
Methodology	This study used the Multiple Criteria Decision-Making methodology (MCDM) to determine a set of reviewers from among the many available options. The weighted sum method was used because the data that have been collected in user profiles are expressed in the same unit. This study produced an experimental result, examining the algorithm with a real collected dataset and mock-up dataset. In total, there were 240 students in the real dataset, and it contained

Accepting Editor Krassie Petrova │ Received: August 16, 2022 │ Revised: November 2, November 13, 2022 │ Accepted: November 17, 2022.

Cite as: Alkhalifa, A., Devlin, M., & Alkhattabi, M. (2022). Matching authors and reviewers in peer assessment based on authors' profiles. *Journal of Information Technology Education: Innovations in Practice, 21,* 135-153. https://doi.org/10.28945/5038

	self-assessment scores, peer scores, and instructor scores for the same assignment. The mock-up dataset created 1000 records for self-assessment scores. The algorithm was evaluated using focus group discussions with 29 programming students and interviews with seven programming instructors.
Contribution	This paper contributes to the field in the following two ways. First, an algorithm using a MCDM methodology was proposed to match authors and reviewers in order to facilitate the peer assessment process. In addition, the algorithm used self-assessment as an initial data source to match users, rather than randomly creating reviewer – author pairs.
Findings	The findings show the accurate results of the algorithm in matching three reviewers for each author. Furthermore, the algorithm was evaluated based on students' and instructors' perspectives. The results are very promising, as they depict a high level of satisfaction for the Balanced Allocation algorithm.
Recommendations for Practitioners	We recommend instructors to consider using the Balanced Allocation algorithm to match students in peer assessments, and consequently to benefit from personalizing peer assessment based on students' needs.
Recommendations for Researchers	Several MCDM methods could be expanded upon, such as the analytic hierarchy process (AHP) if different attributes are collected, or the artificial neural network (ANN) if fuzzy data is available in the user profile. Each method is suitable for special cases depending on the data available for decision-making.
Impact on Society	Suitable pairing in peer assessment would increase the credibility of the peer assessment process and encourage students' engagement in peer assessments.
Future Research	The Balanced Allocation algorithm could be applied using a single group, and a peer assessment with random matching with another group may also be conducted, followed by performing a t-test to determine the impact of matching on students' performances in the peer assessment activity.

Keywords	peer assessment, peer review, matching peers, adaptive peer assessment

INTRODUCTION

Peer assessment involves students reviewing each other's performance against a set of criteria prepared by an instructor or agreed upon by the instructor and the learners. Topping (1998, p. 250) describes peer assessment as considering *"the amount, level, value, worth, quality or success of learning of peers of similar status."* Peer assessment engages students in two roles: "author" and "reviewer". As reviewers, students review their peers' work and contribute feedback; as authors, students receive, read and act upon reviewers' feedback to enhance their own work (L. Li & Gao, 2016). Peer assessment is assumed to lead to more meaningful, in-depth learning because it engages students in the learning process and generates deep learning (Lynch et al., 2012), critical thinking (H. Li et al., 2016), and increased academic performance (Double et al., 2020). Thus, peer assessment is a valuable approach in the learning process.

However, many scholars have highlighted barriers to the use of peer assessment (Indriasari et al., 2020; Zlabkova et al., 2021). One of these barriers is the low learning engagement of students. Some students mentioned that they doubt the efficacy of the peer feedback they receive, because they think that not all peers can provide useful feedback (J. Li et al., 2016). Students in the study by Alkhalifa and Devlin (2021) revealed this, clarifying that suitable pairing in peer assessment could increase their engagement. Anaya et al. (2019) and Capuano and Caballé (2015) observed that many studies ignore the selection of reviewers in peer assessment, as reviewers are determined randomly or manually by experts. Random or manual matching is not viable, especially in online environments which involve a large number of students. Moreover, the social interactions of peer assessments (e.g., characteristics of peers) have still not been addressed in depth (Anaya et al., 2019; van Gennip et al., 2009), despite the fact that social interactions between peers play an important role in peer assessment (Bouguessa & Romdhane, 2015). Thus, this study examined such issues in the context of addressing authors' doubts about the value of the feedback received in peer assessment.

Feedback from multiple peers with different perspectives can sensitize students, make them aware of themselves as authors, and enhance their work quality (Nicol et al., 2014; Topping, 1998). Thus, it is important to include reviewers with different perspectives in each matching group. Social constructivism theory supports differentiation in abilities in peer assessment (L. Li & Gao, 2016). Therefore, the hypothesis in this paper was that matching between authors and reviewers would improve the engagement of peer assessment and provide students with better-quality feedback. The major research goal of this paper was to develop a tool that automatically matches authors and reviewers based on how difficult the author perceived their assignment to be. Authors who submit their work specify how difficult the assignment was, and the algorithm then matches the selected author to a group of reviewers who may meet the author's needs in regard to the selected assignment. The present study outlines a literature review of matching in peer assessment, as well as an algorithm that can determine pairing between peers. Subsequently, a description is provided regarding the development of an algorithm that matches authors and groups of reviewers in order for authors' work to be reviewed. Additionally, the methodology and the experimental results of the developed algorithm, with real datasets and mock-up datasets, are presented. Finally, an evaluation of the algorithm—conducted by collecting the students' and instructors' perspectives—is provided.

LITERATURE REVIEW

BARRIERS TO USING PEER ASSESSMENT

There are several barriers to the use of peer assessment that must be addressed in a successful peer assessment process. The most common barriers are low engagement (Adachi et al., 2018) and low review quality (J. Li et al., 2016). Some students' justification for not engaging in peer assessment is that they do not consider it to be a helpful learning experience (Indriasari et al., 2020). Further, the reviews produced by some students were not consistently trustworthy; there were, for example, instances of inaccurate assessments or low-quality feedback (Alkhalifa & Devlin, 2021). Turner et al. (2008) found meaningless or unhelpful comments in some student review reports. Thus, authors may doubt the reviewers' comments, meaning the peer assessment strategy ultimately negatively influences the learning process (J. Li et al., 2016). The quality of the feedback produced by students can clearly differ based on their abilities (Alkhalifa & Devlin, 2021); since there are differences between the students' abilities in learning, there will certainly be differences in the students' abilities in terms of assessment. Teachers should therefore adapt peer assessment to be based on students' abilities in order to effectively achieve the benefits of peer assessment activity. In addition, the use of multiple reviewers may increase the reliability of the assessment (Indriasari et al., 2020). For instance, assigning more than one reviewer to each author can help students, as they can receive various opinions about their work, so if one author offers poor feedback, another author might be more constructive. In consideration of these matters, this study seeks to find a solution that can improve the quality of review feedback.

MATCHING IN PEER ASSESSMENT

Students have often doubted the effectiveness of peer feedback because they think that not all reviewers can provide them with adequate and credible feedback (Alkhalifa & Devlin, 2021; Kaufman & Schunn, 2011). Moreover, a study conducted by Patchan and Schunn (2016) asked students whether they felt that receiving peer assessment was useful; the students stated that it depends on how knowledgeable their peers are. Students in a study by Alkhalifa and Devlin (2021) concluded that suitable pairing in peer assessment could increase the credibility and engagement of peer assessment, and their receptivity to peer feedback. Since peer assessment is affected by the reviewers' ability (Patchan & Schunn, 2016), more information about the possible kinds of feedback received by learners, and how this feedback differs by ability, is needed. For instance, if an author receives feedback from a reviewer who is highly knowledgeable, the author may receive a significant amount of critical feedback that describes issues and suggests solutions (Patchan & Schunn, 2016). Since a proficient reviewer has better critical skills, they will be able to discover the weaknesses in their peer's text. In contrast, if the author receives feedback from a reviewer who has limited knowledge of the subject of study, the author may receive less critical feedback, which may not sufficiently describe issues or suggest solutions (Patchan & Schunn, 2016). Since a non-proficient reviewer has inferior reviewing skills, they may not be able to make suggestions, but may still offer praise for high-quality work. Another study (J. Li et al., 2016) underlined this notion by proposing a method that could calculate students' assessment ability value, then improve the user allocation algorithm in peer assessment. In order to enhance the credibility of peer assessment, and show peers comments of better quality, creating peer assessment groups with different abilities seemed to be the most beneficial approach.

Topping (1998) theorized peer assessment into social development theory. One of the key concepts of social development theory is the zone of proximal development (ZPD), defined as *"the distance between the actual development level as determined by independent problem-solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with a more capable peer"* (Vygotsky, 1980, p. 86). Vygotsky (1980) stated that a more knowledgeable other is an important aspect of ZPD. Thus, given Vygotsky's focus on individual differences between learners, the present study involved building a matching technique based on ZPD theory to ensure that peer learning can operate at its maximum effectiveness.

Since students are the main focus of peer assessment, they should determine the key variables that the matching process is built upon, to ensure students' engagement in and satisfaction with peer assessment. Building on this, many students, as stated in Alkhalifa and Devlin's (2021) study, suggested an author–reviewers matching technique based on the author's perceived difficulty. Using the self-assessment method, the author specifies how difficult the assignment was, then the algorithm should pair the author to reviewers who did not have difficulty in completing the same task. Therefore, this study was based on the task difficulty level, as a means for assigning suitable reviewers. Some students in Alkhalifa and Devlin's (2021) study also showed concern about whose responsibility it should be to determine the value of the difficulty level: the author or the reviewer. They concluded that both self-assessment by the author and peer assessment by reviewers, with different weighting assigned to each aspect, should be sources for determining task difficulty (Alkhalifa & Devlin, 2021). As a result, selecting suitable reviewers was categorized as a problem of multiple-criteria decision making.

Multiple-Criteria Decision Making

Multiple-criteria decision making (MCDM) is a computational and mathematical algorithm that is commonly used in operations research to assist in the *"subjective evaluation of performance criteria"* via decision makers (Mardani et al., 2015, p. 516). MCDM methods seem suitable for use in author–reviewer matching, allowing the selection of suitable reviewers from the many options available, because the evaluation in the assessment process is based on users' subjective opinions. The MCDM process contains the following main elements: a set of criteria, a structure of preferences, a set of alternatives, and performance values (Mardani et al., 2015). MCDM includes a number of methods that can be implemented based on the available collected data. The weighted sum model (WSM) is the most commonly used approach, particularly in single-dimension problems (Triantaphyllou, 2000), and it is easy to understand and utilize. WSM has been selected in this study because the data that can be collected and stored in the user profiles are expressed in precisely the same units (e.g., peer assessment scores, self-assessment scores). A given MCDM problem has m users' profile alternatives and n

decision criteria: self, peer, and instructor scores. All criteria are positive, given that the higher the values, the better they are considered to be. Therefore, in Equation (1), w_j indicates the relative weight of the importance of the specific criterion C_j, and a_{ij} indicates the performance value of alternative A_i once it is assessed in terms of criterion C_j. The overall importance (i.e., when all the criteria are judged simultaneously) of alternative A_i is indicated as $A_i^{WSM\text{-}score}$. Thus, the top alternative is the one that is, in the maximization case (Triantaphyllou, 2000), defined as follows.

$$A_i^{\ WSM\text{-}score} = \sum_{j=1}^{n} w_j a_{ij} \sum_{j=1}^{n} w_j a_{ij},\ for\ i = 1, 2, 3, \ldots\ m. \quad (1)$$

Several techniques could be used to match users in peer assessment and improve the credibility of the process. For example, fuzzy classification can be used to group users with the same characteristics into a fuzzy set (Ghorbani & Montazer, 2011), with the use of genetic algorithms to find the best fitting set of parameters for pairs. However, fuzzy classification often starts with random matching, before pairing users based on their profiles. This study selected MCDM because the alternatives and criteria are clearly specified, so the MCDM process starts matching from the first round of peer assessment.

BALANCED ALLOCATION ALGORITHM

In a generalized peer assessment model, a set of students assess a specific task for an author. The basic tasks in peer assessment systems (e.g., PeerScholar [Collimore et al., 2015] and PeerGrade [Sharma & Potey, 2018]) include self-assessment and peer assessment, followed by the official inspection and grading of the assignment by an instructor. Therefore, the Balanced Allocation algorithm is able to collect the scores of a specific assignment according to three aspects—self-assessment score, peer assessment score, and instructor score. The algorithm collects the available scores and then assigns the task difficulty level for each author. Difficulty means a reduction in ability to comprehend and solve some aspects of a specific task. Therefore, the author requires feedback from a person who has higher ability, higher skills, or better understanding than the author with regard to a specific task. The authors who submit their work determine, using self-assessment, how difficult their assignment was, and, based on the resultant score, students are organized into two categories: students with difficulties and students without difficulties. Then, a set of reviewers, who may meet the author's needs in regard to the selected assignment, is assigned to each author, taking into account the balance of all matching groups for each peer assessment process in terms of the number of reviewers and ability levels. Figure 1 shows the architecture of the algorithm.

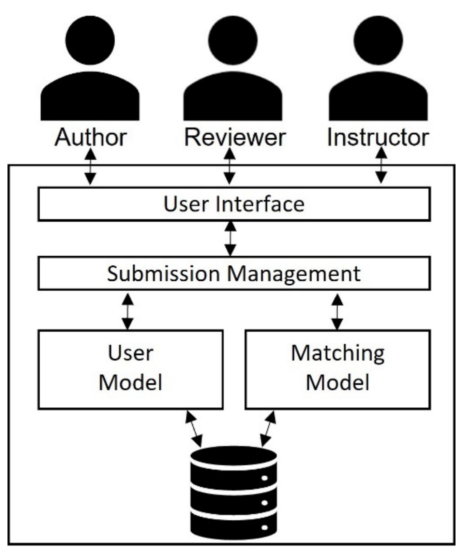


Figure 1: Architecture of the Balanced Allocation algorithm.

Figure 1 shows all possible stakeholders who can assess any specific task through the user interface of a peer assessment system. Author–reviewer matching can be modelled as a scenario consisting of the following steps:

1. The author works on a given task and performs a self-assessment. The author submits the resulting document, which comprises their self-assessment and their solution for the task. The system calculates the author's score in the self-assessment, and keeps the score in the user's profile the first time the system runs (The user profiles only include self-assessment scores in the first round).
2. Based on the user profiles, the students are divided into two groups based on their self-assessment scores: the proficient group who did not experience task difficulty (the top 50% of students) and the non-proficient group who experienced task difficulty (the bottom 50%). The students are divided into two equal groups to avoid a lack of proficient students, especially with a large dataset. Thus, the system decides the task difficulty level of the author, based on their self-assessment, and stores it in the corresponding user profile.
3. Each submission is assigned to a set of reviewers based on the user profiles outlined in the previous step. When assigning reviewers to students, four pairings are possible: 1) a non-proficient student reviews the work of another student with similar abilities, which is a case that should be avoided; 2) a proficient student assesses high-quality work; 3) a non-proficient student assesses work created by a proficient student; 4) and a proficient student assesses the work of a non-proficient student. For the purpose of this study, a non-proficient student's work should be reviewed by at least two proficient students, and the work of a proficient student should be reviewed by at least one proficient student. Consequently, proficient students are able to benefit from reviewing by at least one person of the same proficiency level, and the comments of non-proficient students should not affect their work. Thus, for all users, there are two proficient students and two non-proficient students; hence, the algorithm achieves balanced allocation in each group.
4. After having assessed peers, each reviewer assesses their peer's work; the system calculates the peer score, and then sends it back to the author profile, to update the user profile based on the new value of the WSM. Thus, the tool uses self-assessment and peer assessment to divide students into two groups, assigning different weightings for each criterion.
5. Once an instructor assesses the students, the tool calculates the total sum of the authors' scores for each user profile based on the WSM. This method does not contribute to student grading, in order to ensure that students provide feedback that is as objective as possible.

Figure 2 is a flowchart diagram that explains the Balanced Allocation algorithm process and provides the reader with a visual representation of what occurs. The algorithm is implemented using the R i386 3.6.2 environment (https://cran.r-project.org/bin/windows/base/old/3.6.2/).

First, the user profiles are constructed; they may consist of data on self-assessment, peer assessment, and instructor assessment. However, these three elements are not always available together; thus, the algorithm can work even if only one of the data categories has been collected. Based on the collected data, the tool divides authors into two equal groups: the top 50% of users are categorized as proficient, and their IDs are saved in the Available Proficient List; the rest of the users are classified as non-proficient, and their IDs are saved in the Available Difficulties List. Upon completing this categorization, the loop of assigning three reviewers to each author begins, because, in accordance with the work by Sung et al. (2010), it is difficult to imagine asking students to conduct more than three peer reviews. If the selected author (*k*) is proficient, the selected author should be removed from the Available Proficient List, because the author cannot review their own work. Afterward, the tool selects different random reviewers from two separate lists—one reviewer from the Available Proficient List and the other two reviewers from the Available Difficulties List.

In contrast, if the author is non-proficient, the selected author should be removed from the Available Difficulties List, then the tool selects different random reviewers from two separate lists—two reviewers from the Available Proficient List and one reviewer from the Available Difficulties List. Subsequently, the IDs of the selected reviewers are added to the Reviewer List, which is a matrix that contains two-dimensional elements. To choose another reviewer for a specific author, the selected reviewer should be removed from the Available Proficient List or the Available Difficulties List, because the reviewer is not allowed to assess the work again. The tool counts the number of reviewers for each author to ensure that the review process is equal for all users; in the present scenario, this includes three reviewers. Thus, no non-proficient student can assess another non-proficient student unless there are already two proficient users listed to assess this assignment. As a result, no students assess themselves, there are no repetitions with reviewers assigned to assess the same task, and there are multiple reviewers for each author.

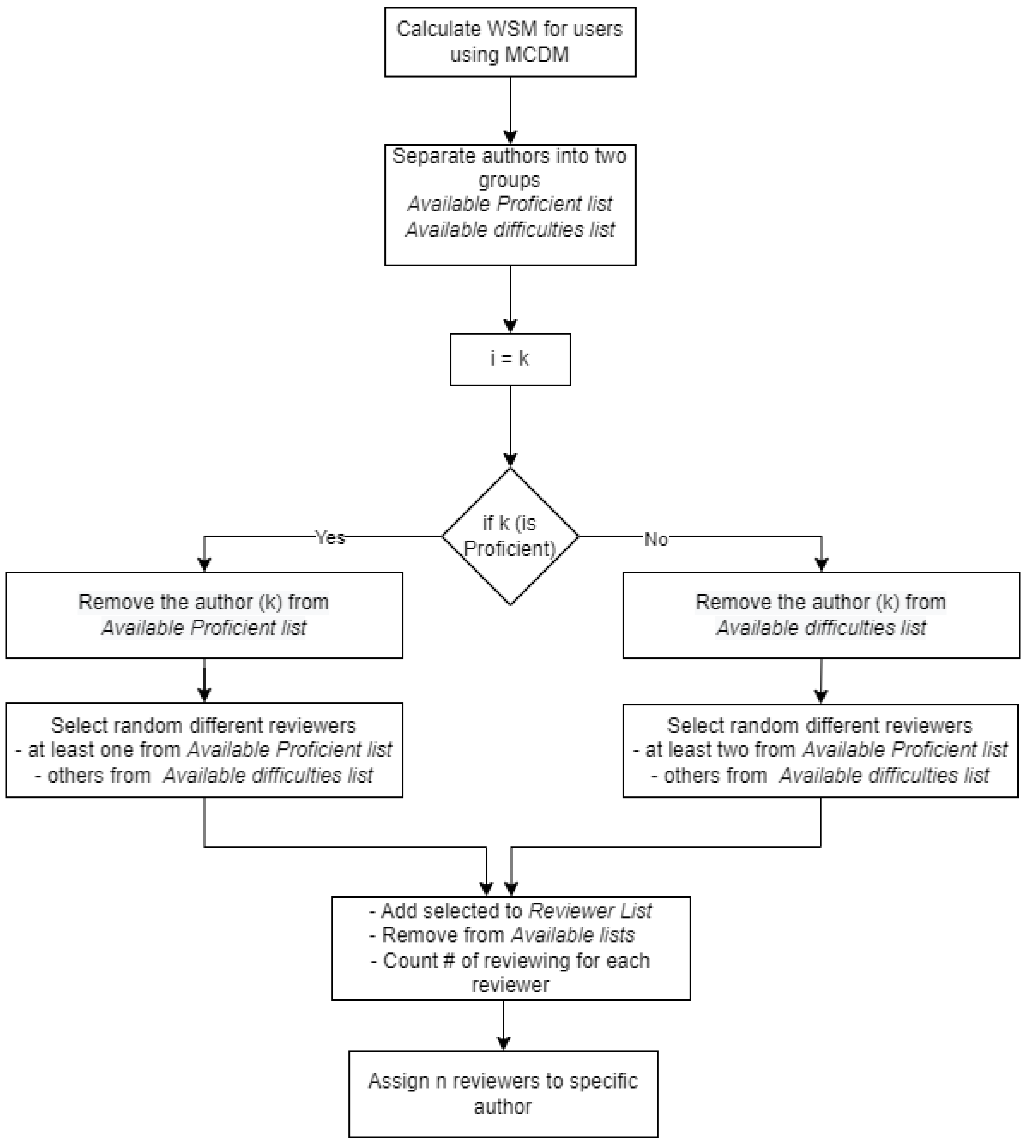


Figure 2: Algorithm Description.

At each stage of the peer assessment process, the algorithm uses the MCDM to estimate the reviewers for each author. Each time the algorithm is used to produce reviewer–author pairs, it updates the user profiles to incorporate recent peer behavior and decide which author should be categorized into which group.

THE PROPOSED METHOD

To evaluate the Balanced Allocation algorithm, actual users' data was selected for use. During the 2005–2008 academic year, Newcastle University, Newcastle upon Tyne, UK was a partner in the Active Learning in Computing (ALiC) project (Devlin, 2015). ALiC is a project focused on increasing the level of student engagement within the computing curriculum, and aims to make their experiences more relevant to industry. In total, there were 240 students in this dataset, with valuable data for the present study. The following information from this dataset was used: summative module marks for all students completing the software engineering course that were graded by their instructor, peer assessment results from the team project, and individual reflective reports completed by the students themselves; therefore, these were classified as self-assessments. Although the dataset is old, it was appropriate for use in this study, because only self-assessment scores, peer scores, and instructor scores for the same assignment were required, and there is a lack of open datasets that combine these variables. These results were then coded in a Microsoft Excel worksheet for the purpose of analysis. The mark data was anonymized, and all records for students who did not finish the module were removed.

Furthermore, to examine the algorithm using a large dataset, the dataset generation website Mockaroo was used (https://www.mockaroo.com/). This online tool generates random data in the form of rows of realistic test data in various formats (e.g., CSV, JSON, SQL, and Excel). The tool creates self-assessment scores only. The researchers did not use the Mockaroo tool to produce random peer and instructor scores, because these three elements are often close if they process the same project or task; thus, such a tool cannot decide random scores that relate to each other. Hence, this experiment focused on producing self-assessment scores. Thus, this algorithm was applied to a real collected dataset and mock-up data for evaluation.

EXPERIMENTAL RESULTS

ALGORITHM IMPLEMENTATION ON A REAL DATASET

Determining the weighting of attributes

The attributes associated with a user profile affected the user categorization results to different extents. As discussed above, based on the user profile, each user was classified as a proficient (having no difficulties in a task) or non-proficient user. Note that this categorization helps to match the authors and reviewers, but it does not affect the students' official scores, as the official scores for the assignments should be decided by the instructor. The weighting associated with an attribute reflects the emphasis to be placed on it; thus, changing the pattern of weightings allocated to various attributes will impact the results of the user's categorization. The following method was used to determine the appropriate weightings for self- and peer assessment. The dataset used for this example contained self-assessment scores, peer scores, and instructor scores for 240 students. The following example (self-score for student i = 60, peer score = 50, and instructor score = 58/100) outlines the method used:

1. All scores were standardized, such that each column was scored out of 10. For example, self-score for student = 6, peer score = 5, and instructor score = 5.8/10.
2. The self-assessment and peer scores were identified, and found to be similar to the instructor scores, given that students' scores for themselves and their peers were subjective. Thus, if the selected score for a specific user was within 0.5 points of the instructor score, researchers decided

that this selected score was similar to the instructor score, while scores outside this range were considered dissimilar. For example, because the self-assessment was 6, it was similar to the instructor score, but the peer assessment score was not similar to the instructor assessment because it was not within 0.5.

3. The number of similar students in terms of self-assessment score and peer score were counted.
4. The percentages of similar results for self-assessment scores and peer scores were calculated as 17% and 37%, respectively. As a result, the weighting of the self-assessment score was 17%, the weighting of peer scores was 37%, and the weighting of the instructor score was 46%.

Method results

In this section, the primary experimental results of the algorithm are presented. Table 1 shows the accuracy results of the algorithm in matching the first, second, and third reviewers for each author. Four random records were selected. Based on the data, the scores of the students were arranged in ascending order, and then the students were divided into two equal groups. The number used to separate the students into two groups—proficient and non-proficient students—was 61.68. Therefore, students who scored higher than this number were classified as proficient students, and the other students were classified as non-proficient students. In Table 1, the total score using the WSM method was calculated using all three attributes (self-assessment, peer, and instructor scores). Authors who were deemed proficient were allocated two non-proficient reviewers and one proficient reviewer. In contrast, non-proficient authors were allocated two proficient reviewers and only one non-proficient reviewer. Table 1, therefore, illustrates two proficient students and two non-proficient students in each row. In addition, the IDs shown here satisfied the algorithm's other conditions: there are three reviewers for each author, no author can assess their own work, no author can be assigned to assess the same task more than once, and no reviewer can assess more than three times during each peer assessment process.

Table 1: Algorithm implementation on a real dataset.

No.	Student's ID Their score	Author	Reviewer1	Reviewer2	Reviewer3
#1	1) ID Score	103 71.7	176 71.4	222 48.4	170 50.4
#2	2) ID Score	139 69.6	136 70.3	213 60.1	141 61.4
#3	3) ID Score	156 58.6	208 81.2	87 64.4	191 56.2
#4	4) ID Score	222 48.4	173 72.9	208 81.2	207 56.3

The scatter plots show the score distributions of the authors and reviewers as a result of the Balanced Allocation algorithm. Figure 3(a) shows that all first reviewers' scores were higher than 61.68 points, whether for proficient or non-proficient authors; this means that for each author, the first reviewer should be proficient. Figure 3(b) illustrates the score distribution of the second reviewer for all authors. There are two clusters: the first cluster represents authors who had scores lower than 61.68 (in this case, the second reviewers should have had scores higher than 61.68); and the second cluster represents authors who had scores higher than 61.68 (in this case, second reviewers should have had scores lower than 61.68). This means that non-proficient authors were matched with proficient reviewers and vice versa. Figure 3(c) shows that all third reviewers' scores are lower than 61.68 points for proficient authors and non-proficient authors, which means that for each author, the third reviewer was non-proficient.

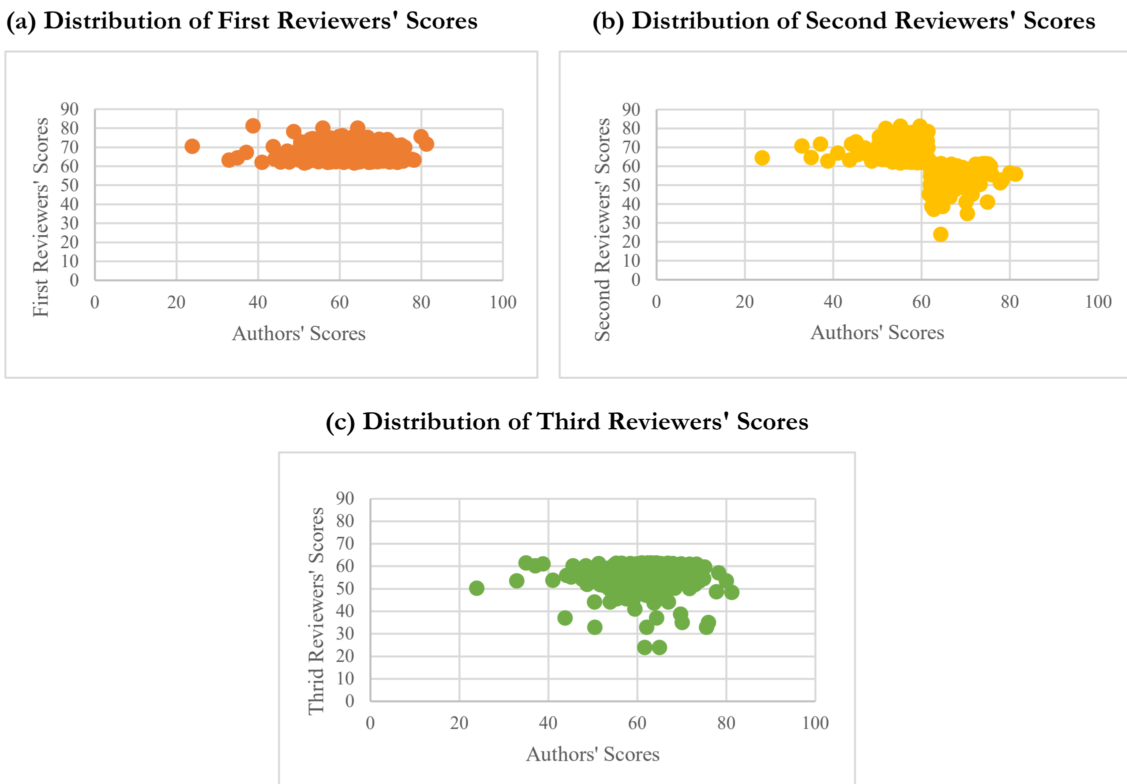


Figure 3: Score distribution between authors and reviewers in a real dataset.

Algorithm Implementation on a Mock-up Dataset

Weighting of attributes

As this dataset was randomly generated, the same weightings were used for all attributes as in the previous case study. The weightings of the self-assessment scores and peer scores were 17% and 37%, respectively, and the weighting of the instructor scores was 46%. As the Mockaroo website randomly generates the self-assessment scores, the total score of the WSM was 17 points, given that the weighting of the self-assessment score was 0.17.

Method results

The algorithm was implemented based on the available dataset, which contained self-assessment scores for 1,000 records. Table 2 shows the results of the dataset. Some author IDs, with matching reviewer IDs and scores, have been randomly selected to display in the table. Based on the generated data, the score that separated students with high and low abilities was 11.22. Therefore, students who had scores higher than this number were classified as proficient students, while others were classified as non-proficient students. As presented in Table 2, if the author had a score lower than 11.22, they were classified as a non-proficient student, and therefore matched with two proficient reviewers and one non-proficient reviewer. By contrast, an author with a score higher than or equal to 11.22 was classified as a proficient student, and was matched with one proficient reviewer and two non-proficient reviewers.

Table 2: Algorithm implementation on a mock-up dataset.

No.	Student's ID Their score	Author	Reviewer1	Reviewer2	Reviewer3
#1	ID Score	1 15.6	497 16.7	604 8.8	913 6.6
#2	ID Score	92 9.4	648 12.9	345 15.0	81 10.2
#3	ID Score	252 14.6	585 16.5	67 10.7	55 5.4
#4	ID Score	593 7.3	367 16.0	29 13.8	97 6.5

The following scatter plots (Figure 4) illustrate the score distributions of the authors and reviewers in the peer assessment as a result of the algorithm using the Mockaroo dataset. Figure 4(a) shows that all the first reviewers' scores were higher than 11.22 points, whether for proficient or non-proficient authors, and were therefore assessed by proficient reviewers (higher than 11.22). Figure 4(b) shows the distributions of the second reviewers based on the authors' scores. As in the previous dataset, there were two clusters: the first cluster included authors who scored lower than 11.22, and the second cluster included authors with scores higher than 11.22. Reviewers for each author belonged to the opposite cluster. The researcher observed that some pairs belonged to the same cluster (both members of the pair were proficient). Figure 4(c) displays the third reviewers' scores.

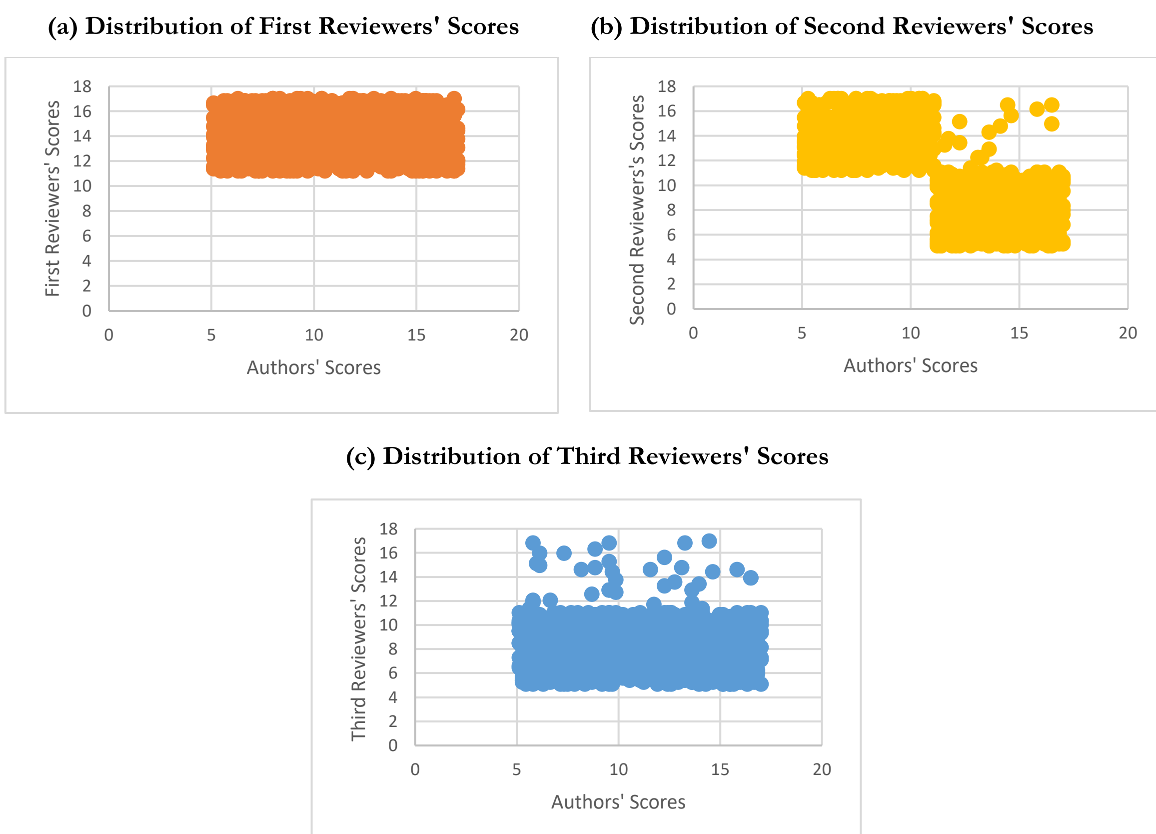


Figure 4: Score distribution between authors and reviewers in a mock-up dataset.

These were below 11.22 points, except in some instances, in which one of the algorithm conditions was selecting (at least) a specific number of proficient reviewers for each case; thus, the algorithm accepted more than two proficient reviewers for non-proficient authors, and more than one proficient reviewer for proficient authors. This situation occurred when all students from the Available Difficulties List were busy, which did not conflict with Vygotsky's theory. The algorithm distributed students based on their abilities, and then paired them in a logical way based on a Vygotsky theory, rather than random selection.

In summary, the algorithm produced the correct outputs for the set of legal inputs. However, it should be indicated that the results may be somewhat limited, as the number of samples in both the real and mock-up datasets was relatively small. Further work is required in order to explore this more extensively, as online environments involve a large number of students.

EVALUATION OF THE ALGORITHM

Given that students in peer assessment activities represent the core factor, the algorithm was evaluated based on students' perspectives. Programming instructors' perspectives were also considered, as they have experience in developing algorithms. The appendix includes the discussion questions for the students and teachers. Ethical approval to conduct this research was granted by Newcastle University (NCL) in the United Kingdom, and Princess Nourah bint Abdulrahman University (PNU) in Saudi Arabia, to which the researchers belong. Participants gave their authorization for discussions to be conducted and recorded.

PROCEDURE OF DATA COLLECTION

The focus groups with students were conducted between 30 September and 16 October 2020. Six online discussions were conducted with 29 undergraduate students who had studied, or were currently studying, computer programming at various Saudi universities. Online focus groups were conducted due to the COVID-19 pandemic. Basic demographic information (study course/level and peer assessment experience) was gathered. Participants belonged to the following departments: computer science (52%), information technology (24%), and software engineering (24%). The participants spanned many knowledge levels: 38% of participants were studying at advanced levels (levels 7 or 8), others (31%) at beginner levels (levels 1, 2, or 3), and 28% at competence levels (levels 4, 5, or 6), with 3% missing data. Most of the students had never engaged in the peer assessment method before. Only 10% of students had used peer assessment informally with their friends; they asked each other to review their assignments before submitting them, without instructors' interventions.

For instructors, structured interviews were conducted online between 15 June and 30 June 2021. Seven programming teachers participated in the interviews—four of them from PNU University and three from NCL University. The instructors occupied various roles—assistant professor (29%), associate professor (57%), and professor (14%). Four programming instructors had used peer assessment in their courses. The other instructors had never used peer assessment before.

All participants were given the consent form, which included the questions form, so that they could outline their ideas. After the study was concluded, the participants were given an online certificate of thanks for their voluntary participation.

DATA ANALYSIS

All the interviews were recorded via Zoom software (https://zoom.us/). Recordings were transcribed using the 'summarized transcript' technique (Baxter et al., 2015) for thematic analysis. After the transcripts were completed, they were revised for accuracy. The coding process came next, including open, axial, and selective coding, with all phases used to analyze the transcript data. Concerning inter-coder reliability, consensus coding was used; an external researcher volunteered to analyze the discussions to review the codes and themes. Then, the researchers and the volunteer researcher discussed and confirmed the codes and themes.

EVALUATION RESULTS

The following three main points were discussed with participants, both students and instructors, to determine the aspects of the algorithm: 'input elements of algorithm', 'process and output of algorithm', and 'tool efficiency'.

Input elements of algorithm

Instructors' Viewpoints: Initially, the interviewer asked participants about the criteria that they recommend for pairing students; the most frequent input was previous knowledge. During the interviews, some instructors indicated their belief that the level of one's previous knowledge influences one's ability to use higher-order cognitive problem-solving skills, such as peer assessment. The interviewer highlighted the impossibility of collecting previous knowledge in some cases—for example, introductory programming courses at the undergraduate level, for which no previous instructor assessment data existed. One participant suggested: *"As a part of that self-assessment, you can ask students if they have a mathematical background, history of learning multiple languages, or experience with problem-solving games."* Further, five instructors agreed on using task difficulty level as an input. As a result, many instructors agreed to use task difficulty level as an input, but some of them suggested adding previous knowledge as another input.

Students' Viewpoints: Eighty-six percent of students agreed on selecting the difficulty level to determine who is assigned to be a suitable reviewer for each author; only 14% disagreed. One student who agreed said: *"If my actual score was not affected, I'm able to determine the difficulty level of the task by myself."* All participants in agreement selected both self-assessment and peer-assessment as methods for assigning the difficulty level of each user, which meant that students could be used as a source of data for user profiles. Then, participants were asked about distribution percentages among the three vectors: self-assessment, peer assessment, and instructor assessment. All participants agreed that self-assessment should not account for more than 20%, while peer assessment and instructor assessment could account for 30–60%. Thus, most of the students supported the use of the task's difficulty level in their user profiles, which can be assigned by themselves and their peers.

Process and output of algorithm

Instructors' Viewpoints: The matching process was explained to the participants, and they were informed that matching would depend on the author's needs, as determined in the Balanced Allocation algorithm. If the author had difficulties in their work, they needed at least two proficient reviewers; if the author had no difficulties in their work, one proficient reviewer out of three reviewers was sufficient. Six of the instructors thought the algorithm included a good balance, and one participant said: *"It is important to have a mix of students who have different ability levels within each group in peer assessment, so that all students in such a group can learn something from that diversity in the group."* Therefore, many instructors were satisfied with student distribution in the algorithm.

Students' Viewpoints: Before explaining the process of matching, most of the participants mentioned that reviewers with a high level of proficiency in programming skills should assess their tasks, which indicated the importance of allocating a proficient reviewer in programming skills in each group. The matching process was explained to the participants too. Most of the students (93%) agreed with the process of this matching and its output. One participant suggested another method for matching: *"It is supposed that the skills be based on our mutual strengths and weaknesses (author and reviewer). I suggest that the reviewer [should] be chosen based on their strengths, which are my weaknesses, to complement each other."* As a result, many students were satisfied with the distribution of reviewers between groups based on the authors' needs.

Efficiency of algorithm

Instructors' Viewpoints: Figure 5 details the instructors' acceptance of the matching technique, their view on the fairness of the matching, its usefulness, and how important this technique would be to students. The instructors found this technique to be highly acceptable (6 out of 7 accepted the algorithm) and useful (all believed in its usefulness). The instructors also thought it was important (only one instructor was neutral) and fair (all agreed on its fairness). None of the instructors held a negative perspective.

Students' Viewpoints: Figure 6 shows the students' viewpoints regarding the algorithm. The students also believed that this technique was acceptable (n = 25/29), useful (n = 27/29), and important (n = 27/29); however, there was lower agreement regarding fairness (n = 19 out of 28). Indeed, there were a few doubts about fairness. The students who did not agree with the fairness of the algorithm thought the proficient authors would not be fair if only one proficient and two non-proficient students assessing their work. They suggested other techniques, for instance, matching based on strengths and weaknesses between the authors and reviewers, or allowing only proficient reviewers to assess peers.

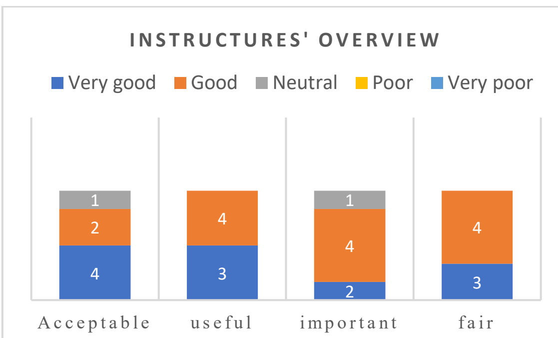


Figure 5: Instructors' overview in matching.

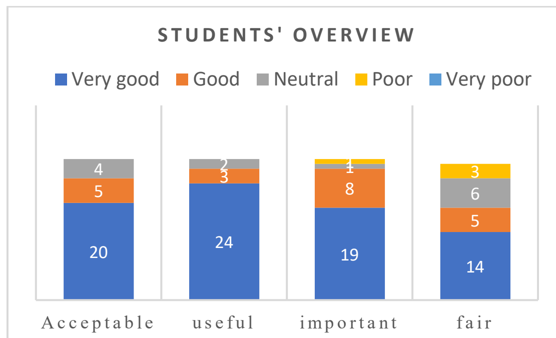


Figure 6: Students' overview in matching.

Ultimately, the evaluation results were very promising. Most of the results depicted a high level of acceptance of the Balanced Allocation algorithm.

DISCUSSION

As has been shown in this study, peer assessment is well suited to author–reviewer pairing; thus, this study has outlined the development of a Balanced Allocation algorithm to match authors and reviewers. This personalized matching is a type of adapting in a learning process, because it carefully assigns a group of reviewers to a particular author to maximize the benefit of peer assessment. The purpose of adaptive learning tools is to provide efficient access to relevant content for the current user, besides providing many features (e.g., just-in-time feedback, pathways, and resources), by creating particular information-based content support (Sarıyalçınkaya et al., 2021). The matching technique used in the current study creates a personalized learning path (by assigning multiple reviewers) and adapts it to the learner's abilities (by considering the task difficulty level). The algorithm developed in this study can then address poor engagement in peer assessment because it meets users' personal needs, and so consequently they could become more engaged by peer assessment.

The pedagogical premise on which the present study is based is the ZPD theory. ZPD is instrumental in understanding "cognitive disorientation", which is characterized in this study by a learner experiencing concern if the assignment presented is too difficult, with the requirement of a more knowledgeable person to help them. When students are in the zone of proximal development, they will improve their knowledge by joining more experienced and competent others (L. Li & Gao, 2016). This thesis has argued that building a tool based on ZPD can assist learners to identify the issues they need to consider, and to find reviewers who can help them with these issues. Other studies have developed adaptive e-learning systems based on the ZPD theory (Imhof et al., 2020; Maravanyika et al., 2017). However, to the best of our knowledge, peer assessments have never included adaptive matching using ZPD. The Balanced Allocation algorithm applied the matching based on ZPD.

Since students are the central part of peer assessment, the input variables of the algorithm that controls the matching process were selected based on students' perspectives, in accordance with Alkhalifa and Devlin's (2021) study. Thus, the algorithm was built based on the task difficulty level. In fact, many adaptive learning systems are based on difficulty levels as a source of personalization information (Kritikou et al., 2008; Tseng et al., 2008). Self-assessment scores were chosen in this study to source the task difficulty level in the user profile. Learners have a realistic sense of their own strengths and weaknesses, and can use knowledge of their own achievements to steer their studies in productive directions (Lew et al., 2010). This study also used peer assessment data to support user profiles; however, peer assessment must be given a stronger weighting than self-assessment. This result does not depart significantly from the findings of Birjandi and Siyyari (2010), who indicated that peer assessment seemed to be more efficient than self-assessment. However, the user profile is currently based on just a few dimensions (e.g., self-, peer, and instructor assessment), so there is a need to expand this set of variables (e.g., friendliness) to improve the sensitivity of the matching provided to learners.

MCDM was used in this study to make decisions with multiple criteria, which needed to be considered together in order to choose between alternatives. The algorithm worked well, and efficiently determined multiple reviewers who could possibly meet the author's needs in regard to the selected assignment. This study assigned reviewers in each assessment process according to the diversity of the ability of the author and reviewers, taking into account the dissimilarity between them. MCDM is usually used to address complex problems with conflicting objectives. It fits a real-life situation in which we have a set of options, and we wish to select the ideal one. MCDM provided a base for selecting and prioritizing reviewers. More generally, MCDM has proven its effectiveness in the field of adaptive learning systems. For example, MCDM has been used to specify tailored learning units for individual students (Chrysafiadi et al., 2019; Kurilovas, 2019). Therefore, MCDM seems to be an effective method for making decisions with regard to selecting reviewers for authors in the peer assessment process.

CONCLUSION

To improve programming students' engagement in peer assessment, this paper suggests a technique to match authors and reviewers in peer assessment. In the study detailed in this paper, this Balanced Allocation algorithm matched authors with sets of reviewers using an MCDM methodology. WSM was used to decide the total scores for all users based on scores from different sources. The real dataset and mock-up dataset were represented to examine the algorithm. The output emphasized the accuracy of the algorithm. Two methods were used to evaluate the algorithm—interviews with programming instructors, and focus groups with programming students. Most of the participants strongly agreed with the need for such a tool to improve their engagement in peer assessment, and to ensure higher-quality feedback. This indicates the importance of considering the selection of reviewers, rather than using random matching, as peer assessment is adapted to students' needs and preferences in order to provide students with high-quality peer feedback. Thus, the results indicate the promising effect of the algorithm based on participants' satisfaction.

This study contributes to the field in the following two ways. First, peer assessment can be personalized based on students' needs by matching authors and reviewers. In addition, the algorithm facilitated the use of self-assessment as an initial data source to match users from the first round, and using MCDM methodology to find suitable reviewers. In future research, several aspects of the present research could be expanded upon. Different types of attributes may be used, such as the analytic hierarchy process (AHP) if different attributes are available, or the artificial neural network (ANN) if fuzzy data is available. The Balanced Allocation algorithm could also be applied using a single group, and a peer assessment with random matching with another group may also be conducted, followed by performing a t-test to determine the impact of matching on students' performances in peer assessment. Moreover, other variables may contribute to increasing the satisfaction of teachers and students with peer feedback, e.g., friendliness, preferences, and social factors. Researchers and practitioners could adopt these variables and conduct more studies to identify their impacts on peer assessment.

REFERENCES

Adachi, C., Tai, J. H., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education*, *43*(2), 294–306. https://doi.org/10.1080/02602938.2017.1339775

Alkhalifa, A., & Devlin, M. (2021, September). Student perspectives of peer assessment in programming courses. *Proceedings of the 2021 Conference on United Kingdom & Ireland Computing Education Research (UKICER 2021)* (pp. Article No.: 8). Glasgow, United Kingdom: ACM. https://doi.org/10.1145/3481282.3481292

Anaya, A. R., Luque, M., Letón, E., & Hernández-del-Olmo, F. (2019). Automatic assignment of reviewers in an online peer assessment task based on social interactions. *Expert Systems*, *36*(4), e12405. https://doi.org/10.1111/exsy.12405

Baxter, K., Courage, C., & Caine, K. (2015). *Understanding your users: A practical guide to user research method* (2nd ed.). Morgan Kaufmann Publishers. https://doi.org/10.1016/C2013-0-13611-2

Birjandi, P., & Siyyari, M. (2010). Self-assessment and peer-assessment: A comparative study of their effect on writing performance and rating accuracy. *Iranian Journal of Applied Linguistics*, *13*(1), 23–45. https://www.sid.ir/EN/VEWSSID/J_pdf/87620100102.pdf

Bouguessa, M., & Romdhane, L. B. (2015). Identifying authorities in online communities. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *6*(3), Article No.: 30. https://doi.org/10.1145/2700481

Capuano, N., & Caballé, S. (2015, November). Towards adaptive peer assessment for MOOCs. *Proceedings of the 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC 2015)* (pp. 64–69). Krakow, Poland: IEEE. https://doi.org/10.1109/3PGCIC.2015.7

Chrysafiadi, K., Troussas, C., Virvou, M., & Sakkopoulos, E. (2019). ICALM: Intelligent mechanism for the creation of dynamically adaptive learning material. *Sensors & Transducers*, *234*(6), 22–29. https://www.sensorsportal.com/HTML/DIGEST/june_2019/Vol_234/P_3088.pdf

Collimore, L., Pare, D. E., & Joordens, S. (2015). SWDYT: So what do you think? Canadian students' attitudes about *peerScholar*, an online peer-assessment tool. *Learning Environments Research*, *18*(1), 33–45. https://doi.org/10.1007/s10984-014-9170-1

Devlin, M. (2015). *The effect of programming competency on success in undergraduate team projects in computing science*. PhD Thesis. Newcastle, United Kingdom: Newcastle University. http://hdl.handle.net/10443/2890

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, *32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Ghorbani, F., & Montazer, G. A. (2011). Learners grouping in e-learning environment using evolutionary fuzzy clustering approach. *International Journal of Information and Communication Technology*, *3*(1), 9–19. https://www.sid.ir/en/VEWSSID/J_pdf/125620110102.pdf

Imhof, C., Bergamin, P., & McGarrity, S. (2020). Implementation of adaptive learning systems: Current state and potential. In P. Isaias, D. G. Sampson, & D. Ifenthaler (Eds.), *Online teaching and learning in higher education* (pp. 93–115). Springer, Cham. https://doi.org/10.1007/978-3-030-48190-2_6

Indriasari, T. D., Luxton-Reilly, A., & Denny, P. (2020). A review of peer code review in higher education. *ACM Transactions on Computing Education*, *20*(3), Article No.: 22. https://doi.org/10.1145/3403935

Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, *39*(3), 387–406. https://doi.org/10.1007/s11251-010-9133-6

Kritikou, Y., Demestichas, P., Adamopoulou, E., Demestichas, K., Theologou, M., & Paradia, M. (2008). User profile modeling in the context of web-based learning management systems. *Journal of Network and Computer Applications*, *31*(4), 603–627. https://doi.org/10.1016/j.jnca.2007.11.006

Kurilovas, E. (2019). Advanced machine learning approaches to personalise learning: Learning analytics and decision making. *Behaviour & Information Technology*, *38*(4), 410–421. https://doi.org/10.1080/0144929X.2018.1539517

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, *35*(2), 135–156. https://doi.org/10.1080/02602930802687737

Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., & Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment and Evaluation in Higher Education*, *41*(2), 245–264. https://doi.org/10.1080/02602938.2014.999746

Li, J., Fu, X., & Yang, Q. (2016, December). Choosing peers, improve the quality of peer assessment. *Proceedings of the 4th International Conference on Machinery, Materials and Information Technology Applications (ICMMITA 2016)* (pp. 1124–1127). Xi'an, China: Atlantis Press. https://doi.org/10.2991/icmmita-16.2016.206

Li, L., & Gao, F. (2016). The effect of peer assessment on project performance of students at different learning levels. *Assessment and Evaluation in Higher Education*, *41*(6), 885–900. https://doi.org/10.1080/02602938.2015.1048185

Lynch, R., McNamara, P. M., & Seery, N. (2012). Promoting deep learning in a teacher education programme through self- and peer-assessment and feedback. *European Journal of Teacher Education*, *35*(2), 179–197. https://doi.org/10.1080/02619768.2011.643396

Maravanyika, M., Dlodlo, N., & Jere, N. (2017, May). An adaptive recommender-system based framework for personalised teaching and learning on e-learning platforms. *Proceedings of the IST-Africa Week Conference (IST-Africa 2017)* (pp. 1–9). Windhoek, Namibia: IEEE. https://doi.org/10.23919/ISTAFRICA.2017.8102297

Mardani, A., Jusoh, A., Nor, K. M. D., Khalifah, Z., Zakwan, N., & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications – A review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istraživanja*, *28*(1), 516–571. https://doi.org/10.1080/1331677X.2015.1075139

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment and Evaluation in Higher Education*, *39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Patchan, M. M., & Schunn, C. D. (2016). Understanding the effects of receiving peer feedback for text revision: Relations between author and reviewer ability. *Journal of Writing Research*, *8*(2), 227–265. https://doi.org/10.17239/jowr-2016.08.02.03

Sarıyalçınkaya, A. D., Karal, H., Altinay, F., & Altinay, Z. (2021). Reflections on adaptive learning analytics: Adaptive learning analytics. In A. Azevedo, J. Azevedo, J. Onohuome Uhomoibhi, & E. Ossiannilsson (Eds.), *Advancing the power of learning analytics and big data in education* (pp. 61-84). IGI Global. https://doi.org/10.4018/978-1-7998-7103-3.ch003

Sharma, D., & Potey, M. (2018, December). Effective learning through peer assessment using peergrade tool. *Proceedings of the Tenth International Conference on Technology for Education (T4E)* (pp. 114–117). Chennai, India IEEE. https://doi.org/10.1109/T4E.2018.00031

Sung, Y. T., Chang, K. E., Chang, T. H., & Yu, W. C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence*, *33*(1), 135–145. https://doi.org/10.1016/j.adolescence.2009.04.004

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Triantaphyllou, E. (2000). *Multi-criteria decision making methods: A comparative study*. Springer. https://doi.org/10.1007/978-1-4757-3157-6

Tseng, J. C. R., Chu, H. C., Hwang, G. J., & Tsai, C. C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers & Education*, *51*(2), 776–786. https://doi.org/10.1016/j.compedu.2007.08.002

Turner, S. A., Quintana-Castillo, R., Pérez-Quiñones, M. A., & Edwards, S. H. (2008). Misunderstandings about object-oriented design: Experiences using code reviews. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education* (pp. 97-101). https://doi.org/10.1145/1352135.1352169

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*(1), 41–54. https://doi.org/10.1016/j.edurev.2008.11.002

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Zlabkova, I., Petr, J., Stuchlikova, I., Rokos, L., & Hospesova, A. (2021). Development of teachers' perspective on formative peer assessment. *International Journal of Science Education*, *43*(3), 428-448. https://doi.org/10.1080/09500693.2020.1713418

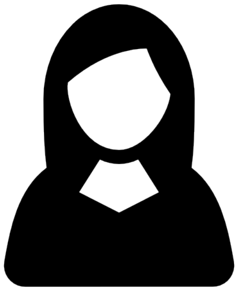
APPENDIX

Discussion Questions for Focus Groups and Interviews

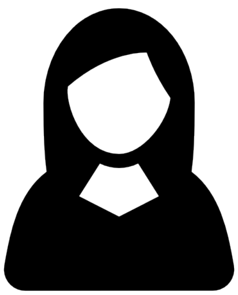
1. What are the criteria you recommend for matching authors and reviewers in peer assessment?
2. Do you think the task difficulty level should determine who gets assigned to be a reviewer for a specific author? Who can assign this ability level?
3. What do you think of the following output of matching: The process of matching depends on an author's need; if the author has difficulties in their work, they need at least two proficient reviewers; if the author has no difficulties in their solution, one proficient reviewer is sufficient.
4. How can the algorithm achieve better outcomes?
5. Please rate the following aspects of the suggested matching process based on your personal view.

	1	2	3	4	5	
Acceptable	□	□	□	□	□	Not Acceptable
Useful	□	□	□	□	□	Not Useful
Important	□	□	□	□	□	Not Important
Fair	□	□	□	□	□	Unfair

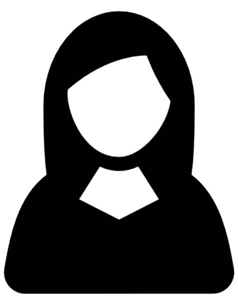
AUTHORS



Amal Alkhalifa is currently working toward a PhD degree in computing science at Newcastle University, Newcastle upon Tyne, UK. She is also working as a lecturer with the Computer Science and Information Technology Department, Applied College, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Her research areas include assessment, adaptive learning, and educational technology.



Marie Devlin is an Associate Professor in computing at the School of Computing, Newcastle University. She is also the Associate Dean (Undergraduate Education), SAgE Faculty of Science, Agriculture & Engineering at Newcastle University. Her research areas include methods for assessing software engineering competency, technology-enhanced learning in computer science education, and assessment and feedback in higher education.



Mona Alkhattabi is an Associate Professor at the College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. She is also the Vice Dean of Information Technology Deanship, Imam Mohammad Ibn Saud Islamic University. Her research areas include e-learning, business intelligence, data mining, the semantic web, and HCI.