Improving Grading Fairness and Transparency with Decentralized Collaborative Peer Assessment

ANONYMOUS AUTHOR(S)*

Computer-assisted collaborative peer grading is a developing growth area in academic evaluation. However, peer assessment often needs help with problems such as the lack of reliability, transparency, fairness, grading speed, and motivation to participate among students. The literature suggests several principles that each partly address the said issues. We propose a novel decentralized approach to academic peer assessment, using blockchain as an underlying technology, to address the principal problems in traditional peer assessment. We also derive design concepts for a modern courseware (CW) application consisting of our method and apply them to implement our approach in a CW called `Blockment`. We test the effectiveness of our method and system by running quantitative and qualitative experiments, proving our claims of improving reliability, transparency, fairness, grading speed, and motivation of grades in peer assessment. The results suggest embedding our method and system in academic courses to improve conventional peer grading methods.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**.

Additional Key Words and Phrases: peer grading, blockchain, courseware

ACM Reference Format:
Anonymous Author(s). 2018. Improving Grading Fairness and Transparency with Decentralized Collaborative Peer Assessment. In . ACM, New York, NY, USA, 23 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Peer assessment (also called peer grading) is an educational setting in which groups of participants provide other students with feedback on the quality of their work (e.g., homework or exam solutions) and assess their peers' effort [66, 87]. In peer assessment, instead of or in addition to instructors and teaching assistants, students participate in the grading process [82]. In a peer assessment scenario, individuals evaluate the value, amount, or success of the products or outcomes of the learning process of peers with similar status to them [82]. Peer assessment is a recognized, unique form of collaborative learning, that has become a well-known pedagogical method in higher education [66].

Peer assessment is widely used in many classrooms worldwide [25], usually incorporated into computer applications by researchers in computer-supported collaborative learning (CSCL) [63, 66]. Peer grading has numerous benefits per the literature, including but not limited to improvement in students' skill acquisition [26], advancement of critical thinking in students [18], positive effect on student attitudes [82], and better learning outcomes [68].

However, the computer-supported peer assessment systems currently in use suffer from some drawbacks, such as low reliability of the scores given by the peers [14, 51], lack of transparency [96], the effect of personal bias [8–10, 38, 64], high grading burden [39], and lack of intense motivation [48, 85]. While the previous works have attempted to find solutions for the problems mentioned above and evaluate them, each has only focused on a few issues. Hence, there is a

Manuscript submitted to ACM

literature gap regarding a transparent, fair, and fast peer assessment framework that addresses the well-known issues of traditional peer grading systems.

In this paper, we provide a decentralized peer assessment method, a viable approach to peer grading, correctly and sufficiently addressing the issues commonplace among traditional peer assessment systems. We build our decentralized system based on a blockchain network since a blockchain core assures us a high level of transparency in the grading process. We summarize the common issues with peer assessment from the literature and address them one-by-one by incorporating specific principles in our decentralized peer assessment system. Moreover, we find the potential problems with using a blockchain-based system in an academic environment and provide feasible solutions.

In addition to providing our novel decentralized peer assessment method, we conducted interviews with students who had previous experience with peer assessment. We used the outcomes of the interviews and the literature to derive design concepts for a user-centric courseware (CW) software with support for peer assessment. Based on the design concepts, we designed and developed `Blockment`, a web-based CW application to help students, instructors, and teaching assistants in a classroom use peer assessment based on our decentralized method in their courses. As a result, in addition to providing our decentralized system, we also state how we should embed it in a user-centric design in real-world pedagogical scenarios.

To evaluate if our decentralized peer assessment method is effective and addresses the common problems it aimed initially to solve, we ran two experiments on a total of 52 students. We asked the students to provide us with their perception of the fairness of the grade and grading speed to analyze them quantitatively, alongside the raw grades to find insights into grade reliability. We also asked them to answer open questions to find insights into the transparency of our method and the motivation of the participants, which we analyzed qualitatively. In addition, we tested the user experience of `Blockment` in a laboratory usability experiment with eleven current or former students in a think-aloud process to find if people see our design as intuitive. The results from the experiments show that our proposed system successfully addresses the significant issues with the previous peer assessment schemes. The contributions of our work are threefold:

- We provide a novel peer assessment system built over blockchain and show how it successfully addresses the common issues among most other peer assessment schemes, collected from the literature (as discussed in Section 2.2).
- We extract the possible issues with using a decentralized system and address them by adding specific principles to our blockchain backbone system.
- We contribute to the current literature on CSCL and human-computer interaction (HCI) by collecting and providing design concepts for a user-centric interface, as well as implementing and evaluating it in a laboratory usability experiment, that operates on our proposed decentralized backbone without exposing the technical difficulties of such a system to the users.

We implement a web application based on our collected design concepts called `Blockment`. Our proposed backbone system and CW design aim to solve the commonplace peer assessment issues. They can be used in various educational environments to improve the transparency and fairness of grades obtained collaboratively in a peer grading scheme. Our evaluation of `Blockment` in our laboratory experiment suggests embedding `Blockment` as a viable peer grading CW software in higher-education courses.

2 RELATED WORK AND CONCEPTUAL BACKGROUND

This research was inspired by the literature around peer assessment and its benefits, computer-supported collaborative learning, previous attempts at addressing the issues of computer-supported peer assessment, and blockchain and its role in education.

2.1 Definition and Benefits of Peer Assessment

Peer assessment is a pedagogical scenario in which students collaboratively provide their peers with qualitative and quantitative feedback on their work and evaluate the effectiveness of their peers' learning process [66, 82, 87]. Peer assessment is different from the traditional methods of grading in the sense that grading was traditionally the sole responsibility of the tutor [22]. Peer assessment is a method for collaborative and social learning, which has recently received attention as a viable frame for learning [32]. While peer assessment is performed mainly by the students, it is considered a "teaching activity" [30]. Providing reviews to peers on their work has been widely used in classrooms since the 1970s and is rising in adoption among educational environments [25, 42, 57, 77]. Peer assessment is conducted in various ways, depending on the course topic and structure, and is not limited to grading papers [25]. For example, it can help facilitate code reviews in computer science courses, evaluate the contribution of team members in course projects, and judge the quality of the documentation provided along with the code in programming projects [25].

Numerous benefits to peer assessment have been explored and discovered in the literature. Although the attitude toward peer assessment differs among student bodies [28], students benefit from and learn by both receiving and providing peer evaluation, which helps them find valuable methods to improve their skills [68, 83, 95]. In a study by Georgouli et al. [26], students expressed satisfaction with collaborative peer evaluation in the class. Also, Cheng et al. [13] have shown that peer assessment exercises lead to a shift in the students' attitudes. Peer assessment is considered beneficial, valid, and reliable, and students positively evaluate the peer assessment procedure as being beneficial [13, 43, 60, 61, 81], especially in a blended learning context [12]. Providing educational reviews to peers is known to develop reflective processes, such as critical thinking [18, 45] or cognitive skills [72]. As such, a classroom is a fitting place for providing suggestions on their work to peers since the subject is fresh and there is a high possibility of in-depth interaction among peers [18].

Moreover, peer assessment leads to effective grading of written material [87], positive formative effects on the attitudes and achievements of students [82], and increased personal motivation due to the active involvement of the students [9]. It allows students to improve their responsibility and autonomy [13], master academic concepts [78], understand the grading process and the assignment content better, and compare and discuss the assignment [9]. As such, peer assessment has been successfully employed in various academic environments, leading to several benefits [72]. It helps build a strong learning community [74] and improves the efficiency of educational procedures under the digitalization of education at schools [65], ultimately leading to addressing the concerns regarding *quality education*, the fourth goal of the United Nations Sustainability Goals [93].

2.2 Computer-Supported Peer Assessment

While peer assessment can be also applied to traditional classrooms without the use of new technologies, computer-supported approaches can be also used to facilitate this process. Computer-supported collaborative Learning (CSCL) is a promising idea to help learning and tutoring progress with the assistance of novel information and communication technologies (ICT) [41]. Numerous computer systems have been implemented to support students in peer assessment, many dating back to the early days of the world wide web. Filippakis et al. [23] have developed a web application to help distribute assignments in peer evaluation scenarios. Gehringer [24] has implemented a peer grader system that encourages careful review of the submissions. Trivedi et al. [83] have designed a web-based system for anonymous peer evaluation and online course management, with a rubric designed and maintained by an instructor. Davies [17] has addressed the issue of plagiarism in computer-supported peer assessment systems. Also, in recent works, researchers have designed and implemented systems for using collective intelligence to support students in sharing knowledge [59]. They also used conversational agents and chatbots to guide students in the peer assessment process [40], and cloud teaching assistant systems in massive open online courses (MOOCs) [89].

However, there are several limitations in the previous peer assessment systems, which researchers have independently tried to address by introducing specific principles. These limitations form the motivation to perform the current research:

2.2.1 Low reliability of the scores. Cheng et al. [14] believed peer assessments are not sufficiently reliable for being used instead of or in addition to conventional teacher assessments. Also, by calculating inter-rater reliability, Magin et al. [51] found that teacher ratings have a significantly higher inter-rater agreement than peer rating. They claim it would require averaging the scores given by up to four peer raters to match the reliability of one teacher.

2.2.2 Lack of transperency. Smith et al. [73] have raised the issue of low transparency in conventional peer assessment methods. They witnessed increased transparency by using a trial marking exercise and intervention by the experiment designers [73]. Perdue et al. [62] examined the role of transparency on peer review quality in MOOCs and found that transparent peer assessment produced higher-quality feedback than other systems.

2.2.3 Effect of personal bias on fairness of the grades. Brindley et al. [9] have discussed the effects of personal bias on the awarded marks. Kulkarni et al. [38] found that giving students feedback on their grading bias boosted subsequent accuracy. On the other hand, removing personal bias through anonymity may lead to a decrease in transparency [62]. In order to characterize the relationships of bias among the students, Chan et al. [10] suggested three probabilistic models for peer assessment. In order to account for the various bias and precision of student assessors, Bradley [8] has investigated the usage of a hierarchical Bayesian model and demonstrated that the model is solid and improves assessment fairness in actual assessments.

2.2.4 High grading burden. According to Kulkarni et al. [39], peer assessment forces students to spend a lot of time grading the work of their peers even while it aids in reflection and exposes them to various viewpoints. They combine peer and automated grading to maintain the integrity of peer assessment and lighten the grading burden. They illustrate how combining peer collaboration with machine learning might enhance learning.

2.2.5 Lack of strong motivation. The issue of student's lack of motivation is brought up by Lu et al. [48] They discover compelling statistical proof that "rating the graders themselves" improves the quality of peer grading. Also, Turner et al. [85] demonstrate the critical necessity for outside motivation to get students involved in peer assessment scenarios.

2.3 Blockchain and Blockchain-based Systems

A blockchain is a growing list of *records* (called *blocks*) linked together using cryptography methods, each containing a hash of the previous block, transaction data, and timestamp [58]. The first application for Blockchain was provided by Nakamoto [58], in which Nakomoto introduced a decentralized transaction ledger to implement the Bitcoin cryptocurrency.

Bitcoin's system, as a well-known example of a blockchain-based system, operates in a decentralized fashion without relying on a trusted third party. Since its introduction, no fraud on the blockchain network has been found [27]. A blockchain network ensures transparency since a transparent log of every transaction is visible to everyone observing the network [4]. It also ensures security, reliability, and data integrity [2]. Blockchain is a new type of infrastructure that has the potential to fundamentally alter the way that people interact, cooperate, organize, and self-identify [21].

In addition to cryptocurrency and Bitcoin, blockchain has been used as a viable backbone in implementing various HCI and user-centric systems [16, 21, 86]. Researchers have used blockchain for trusted timestamping [31], proof of patents and copyright [79], digitalized land registration [88], agriculture [3, 52], and efficiently managing water use under climate changes [44], to name a few.

Blockchain has also been used in educational and pedagogical settings [2, 67]. Researchers have implemented blockchain-based solutions in numerous scenarios, including but not limited to managing verifiable educational and course certificates [5, 29]. They have evaluated students' professional ability through cooperative systems [46, 47], securing a collaborative learning environment [6, 7, 99], fees and credits transfer [33, 91], enhancing students' interactions in e-learning [99], and supporting lifelong learning [54].

3 OUR PROPOSED METHOD

In this section, we first describe the reason behind using a blockchain platform as a base for our method. Then, we describe an overview of our proposed method and details of the assessment procedure.

3.1 Why Blockchain?

To solve the issues introduced in Section 2.2, we examined various methods that can be implemented to improve peer assessment. We claim that using blockchain-based systems can be a viable solution to the issues mentioned above. We have identified numerous advantages [16, 86] of using blockchain-based systems in peer assessment, each solving one of the issues in Section 2.2: Blockchain-based systems are reliable, as no single entity can alter the data in the system [75]. They are also transparent since all the data in a blockchain network is visible to everyone observing it [4]. Blockchain-based systems also decrease the effect of personal bias since they can remain anonymous [1, 90, 98]. Blockchain-based systems also decrease the grading burden since no centralized entity is needed to grade the works or collect the students' grades [97]. Finally, blockchain-based systems can increase the motivation of students, since they can be gamified in the sense that students can earn rewards for their work (e.g. higher grades or grading power) [36, 71]. As a result, using a blockchain platform as the core of our method ensures that our system is reliable, transparent, unbiased, with a low burden, and motivating. Such a system has the potential to satisfy both students and teachers of a given course at the same time.

3.2 Overview of Our Method

In our method, students start by signing up for the courses. The record of each student's registration in the course will be put on a blockchain. As the students download or watch lectures, the records will be updated accordingly by inserting new log records in the blockchain. Meanwhile, homework handouts are designed and released collaboratively, and the students submit their answers to the decentralized network. Each homework solution will be assessed by a sample of people including three user roles: the instructor, the teaching assistants, and the other students taking the same course (*classmates*)[1]. After completing a course, the students will get a certificate if they have eligible grades, and the certificate is pushed to the blockchain, making it verifiable, immutable, and transparent. In our proposed process, *validator nodes* (e.g., educational institutions) provide the users with the available data and prevent the need for storing the whole data on every users' device while preserving the privacy, distribution, and transparency. Our method supports registered students through the whole learning process and will reward them for their effort in the course. In our method, enrolled students learn the material as the course moves forward, and throughout the progress of the course, they design new homework handouts and ask fellow students to send their answers. After each class member solved the problem set given to them, which other learners developed, they begin to assess their fellow students' solutions.

3.3 Details of the Assessment Procedure

The assessment procedure in our method works as follows:

(1) Students submit their homework anonymously to validator nodes of their choice. The submission is encrypted with public keys of those validator nodes.
(2) Each validator node confirms the receipt of the submission by sending the solution to an encrypted challenge.
(3) The students, the teacher, and the teaching assistants *compete* to assess the submitted homework. After assessing the homework, they release the papers, the assessment details, and the given grade, which are gathered into a block and added to the blockchain. Their knowledge score in the system and the assessment time determine the priority and impact of that block and the rewards they receive. The students get a small amount (one point) of score for assessing each paper. At least three students should assess each homework.
(4) The students act as miners, which means they check the assessment details of papers and add the verified assessments to the blockchain. Each block consists of the homework/exam questions, responses, and grading results. The miners get a score (similar to Bitcoin in the Bitcoin blockchain [58]) for verifying the grades. In our method, the winner is not who possesses better hardware (similar to the case in the Bitcoin blockchain [58]), but who has a better knowledge of the course and is more proficient at assessing other solutions. This makes our method in line with improving educational outcomes. To ensure this, every student gives a half-point of score to the student that assessed their paper, so that the blocks associated with students with higher assessment scores have a higher chance of being accepted.

In Figure 1, we have provided a simple diagram of how our proposed method works.

[1]The details of the assessment procedure are described in the rest of this section.

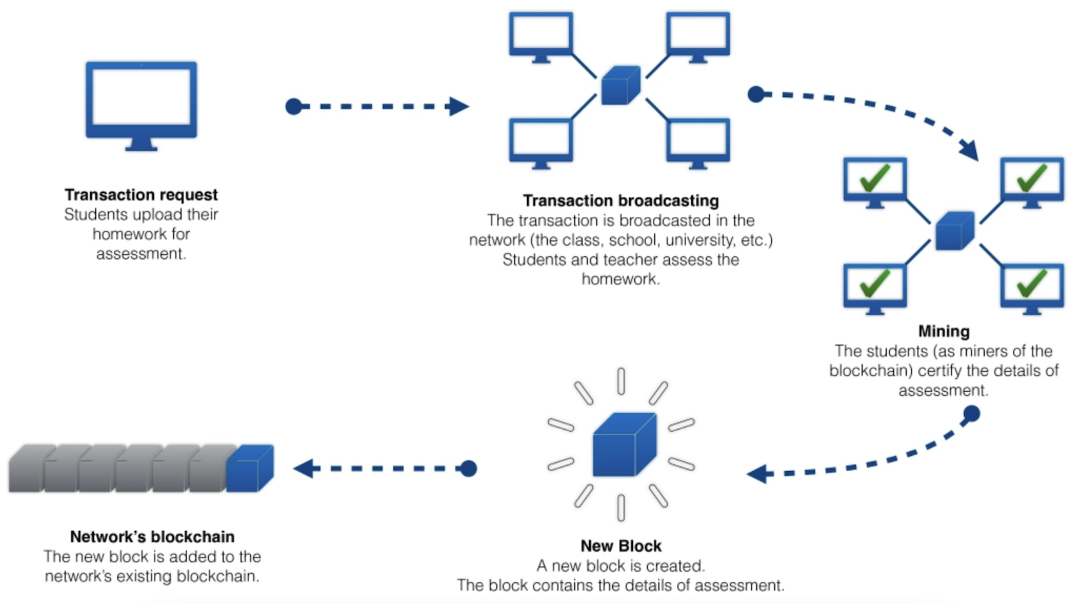

Fig. 1. A simple diagram of our proposed method. Students upload their homework, which is broadcasted to the network. Others assess the homework and help in certifying the assessments, and the new grade is added as a block to the blockchain.

To incentivize the users, we propose a metric called the *reputation score*. This score is used by the validator nodes to prioritize submissions and assessments and is also used to assess the credibility of assessments.

New users have a base reputation score of 10. When a work submitted by a user is scored by other users, the user gains a reputation score equal to the weighted average of the given scores, where the weights used for averaging are the reviewers' reputation scores. To have fair conditions, the given scores are normalized before averaging so that the possible score range would be from -10 to $+10$. For example, if the scores are out of 100 and the minimum possible score is 0, they are first divided by 5 and then reduced by 10 before averaging.

In addition, if an assessment receives a review, the writer of the assessment will gain a reputation score equal to the score they receive, normalized similar to the example above so that the possible score range would be from -5 to $+5$.

We also propose a native token called *Stamina*, to incentivize validating and prevent spamming. Users are charged one Stamina for each data submission and are given two Stamina for each assessment. The validator nodes can not issue Stamina for themselves directly, but they manage its distribution to the users with the said rule.

If users run out of Stamina and can not get any by assessing, they may request the validators to give them 5 Stamina, with the condition that a week must have passed from their last request. New users do not have any Stamina to start with, but they can request to recevie it as explained before. Stamina can also be issued by trading in reputation score, where one reputation score gives 2 Stamina. Unlike reputation score, Stamina may be transferred willingly to another user.

3.4 Details on System Architecture

To provide users with authenticity, availability, and privacy, we propose a distributed network with validator nodes and user roles including *student*, *teacher*, and *teaching assistant*. The main distributed database is propagated through the validator node network and normal users can access the whole database by interaction with the validator nodes, without needing to store it completely on their own devices.

Each user is defined with a (public key, private key) pair. The public key and the private key are effectively the user's username and password, respectively. The distribution of the public key is independant of our proposed system and users can share their public key by any means they desire.

New users are considered students by default. For a user to be recognized as a teacher or teaching assistant for a specific course, they need to be given the role by an existing privileged user in the course. The user who creates a course is its sole teacher upon creation, and they can assign the privileged roles to other users afterwards.

We propose to use two chains, the *mainnet* and the *briefnet*. The mainnet is the main chain and contains the whole system data including the roles, the files, the courses, and the scores. Every validator node keeps a copy of the mainnet and changes are propagated through the network. Users can retrieve copies of specific parts or the whole mainnet from the nodes at will. The briefnet contains only essential data, including data about the courses and the material available for scoring. Every user and validator node keeps a copy of the briefnet.

A typical block of the mainnet contains the *block number*, the *submission ID*, the *user IDs*, the *data*, the *encrypted data*, the *data description*, the *transaction data*, the *signature count*, the public keys for the *signer validator nodes*, a *Nonce*, and the *Hash of previous block*.

A typical block of the briefnet contains the *block number*, a *submission ID*, the *user IDs*, the *data*, the *data description*, a *nonce*, an the *hash of the previous block*. It has a similar structure to the mainnet, but eliminates parts such as encrypted data. Blocks from the mainnet and the briefnet are associated together using the submission ID.

The identity of a validator node is also defined with a (public key, private key) pair, just like normal users. The distribution of validator node public keys is also independant of our proposed system and nodes can broadcast and share their public keys by any means they desire.

With the proposed structure, we want to achieve privacy for the users and prevent unauthorized access (e.g., students accessing the answers from other students before submitting a solution themselves) while preserving the availability and transparency. Sensitive data such as files are encrypted before being stored.

For increased security, the *encrypted data* part of the mainnet blocks is encrypted with the public key of the validator node the mainnet is stored in. We call this a *signature* from the validator node. Depending on the use case, users can set the *signature count* to a number higher than one. In any case, the public keys of all *signer validator nodes* will be included in the respective block part, and the collaboration of all involved nodes is needed to decrypt the data.

To submit a piece of data, the users must select a signature count N for their proposed block. They then select N validator nodes of their choice and create a *challenge* (as described below) for each. After confirmation with the nodes, the user consecutively encrypts the sensitive data with all of the nodes' public keys and sends the request to the nodes.

A *challenge* is created, and subsequently verified, through the following steps.

(1) The user produces a nonce (a large random number), and encrypts it with their own and the destination validator node's public keys.
(2) The user sends the encrypted nonce to the node.
(3) The node decrypts the encrypted nonce with its own private key and sends it back to the user.

(4) The user decrypts the received response and compares it with the nonce. If identical, the challenge is pronounced successful; otherwise, it is rejected.

After receiving a valid new request, the validator node creates a new block and adds it to its own mainnet. It then sends a verification message encrypted with the user's public key to the user. It then selects some other validator nodes and for each, it generates a new request by decrypting the encrypted data with its own private key and then encrypting it with the destination node's public key. If the data is not new, the request will be ignored.

This architecture causes the mainnet on different nodes to be different and unique, but the redundancy and the fact that the mainnets on all nodes contain the same data, but with different keys, prevents mutability of blocks created in the past. Duplicates are also recognized using a *submission ID* that is shared with the respective block from the briefnet.

After receiving the confirmation from the respective validator nodes, a block generated for the submission is added to the briefnet by both the user and the node and the change is then propagated through the network. When a party receives a change request to the briefnet, they validate it with other users and validator nodes before accepting it.

The transactions of reputation score and Stamina are also submitted to the mainnet just like any other kind of data. If the transaction is dependant on another submission (e.g., submitting a solution, creating a course, or assessing other submissions), the data is written in the transaction *data* section in the same block as the main submission. Otherwise, in the case of transmitting Stamina, a new submission is created for the transaction and the details are written in the *data* section and the affected user in the *user IDs* section.

The Stamina balance and reputation score for each user can be determined by either querying validator nodes, or by retrieving a copy of the whole mainnet from a validator node and computing the balances by taking into account the transactions with the desired user in the *user IDs* section.

3.5 Example of Our Method in Practice

For example, imagine Paul and Carole are the only two students enrolled in a course (a simplified situation). The course consists of Merlin (the instructor) and Arthur (the only TA). At first, before the enrollment in the course started, Merlin had prepared selected readings, as well as video lectures (in case the class is a flipped classroom or a MOOC), and he had uploaded them to the network; the material might have also been checked by Arthur (the TA). Then, after the registration start date, Paul and Carole register for the course. Their records get updated (by inserting a new record with updated details of the student's profile, indicating the passed courses and necessary certificates) in the blockchain, ensuring transparency and proof-checking one's skills by any observer easily (just as with other blockchain systems [4]).

After the enrollment of Paul and Carole (the students) in the course and reaching certain checkpoints in the lectures and/or readings set by Merlin or Arthur, they will be asked to design problem sets based on the lecture contents and also similar to some example initial problem sets possibly given by Merlin and/or Arthur to them, to guide them on how to design problems for that specific topic.

After the designed problem sets are released, a selected number of questions (determined by Merlin or Arthur at the beginning) will be assigned to each student without them knowing which problem belongs to which person (the problem sets should be submitted in an anonymized way). Each student will answer the problems assigned to them, and the problem set they have designed and submitted before. Preparing the answer to the questions designed by the student is necessary because it will later act as a solution key for grading the answers.

After the answers are collected, the answers, along with the solutions provided by the designer of the problem, are assigned randomly to the registered students to grade, as explained earlier. When grading is finished, all solutions and the assigned grades are inserted into the blockchain, ensuring the immutability and transparency of the grading results. The blocks associated with students who have higher assessment scores can have a higher chance of being accepted. The final certificate issued to each student is easily verifiable by running a computer program (or observing by an individual) over each student's records in the publicly-available blockchain of our method.

4 DESIGN OF BLOCKMENT

In Section 3 we claimed that our proposed method satisfies our desired characteristics of a peer-grading scheme. Nevertheless, our method also needs to be included in a user-centric CW software in order to be able to be used effectively by the students, teachers, and teaching assistants.

In this section, we present the design of `Blockment`, our blockchain-based peer CW software for peer grading, indicated to be used in university and specifically undergraduate courses. `Blockment` acts as a user-centric wrapper around our proposed model and helps to enforce necessary policies and rules for running our method based on requirements from the literature. We also designed a prototype of `Blockment` to test the feasibility of our proposed design.

4.1 Design Issues in the Literature

Previous works have identified critical issues in the design of blockchain-based platforms and computer-supported peer grading systems. This section discusses the issues identified in the literature and how we address them in our design. To find the issues, we reviewed the previous literature by searching for the keywords "peer grading issues," "grading challenges," "challenges in blockchain systems," and their combinations in Google Scholar. We analyzed 11 papers and identified four clusters relevant to our work, which we consider them all in designing `Blockment`:

- **A) Self-Grading [70, 76]:** `Blockment` ensures that the person who assigns the grade of a particular handout is not the person who graded it in the first place. Even though the users participate anonymously, it is evident that students remember whether a specific handout is theirs. While students feel more motivated to learn in self-grading scenarios, grade inflation may occur in such settings.
- **B) Grade Verification [36, 71]:** To increase transparency and verification validity in a blockchain network, `Blockment` rates each user and validates their new assessments by a reputation score. `Blockment` rewards scores to the users as an incentive for checking and correcting to address the need for more specific checkers and correctors. The blockchain includes the final reputation of each user as part of their certificate. We have previously discussed the details of the reputation score in Section 3.3.
- **C) Scalability [15, 50, 69, 92]:** Peer grading systems are usually designed for small groups of students, which is not scalable to large classes. In addition, the number of students willing to grade the homework is usually less than the number of students who submit their homework. Our blockchain-based system improves the scalability of peer grading by using a distributed network of users who all participate in the grading process and are assigned handouts to grade by our system.
- **D) Usability [34, 37, 80]:** To use a blockchain-based method for peer grading, the users should be familiar with blockchain technology, which is a potential problem for the students. Few works have considered user-centric interfaces when developing their blockchain-based applications, which leads to usability issues. Consequently,

our user-centric platform's user experience does not show any of the system's blockchain underpinnings and technical sides on the surface. It resembles traditional CW software, making it usable and suitable for people who are not necessarily familiar with blockchain technology.

4.2 User Interface of Blockment

Based on the requirements from the literature, we designed `Blockment`, our user-centric platform for peer grading. `Blockment` is a CW software that helps students submit and grade each other's homework. It is a user-centric wrapper around our proposed model, which includes specific policies for running our method from the literature (as described in Section 4.1). `Blockment` is a responsive web app that can be opened and used on various screen sizes and devices (apart from smartphones). The back-end API of `Blockment` is developed in Python using the Flask[2] framework, and the front-end is developed in JavaScript, HTML, and CSS.

The front-end design resembles traditional CW software, where users can enroll in a course, submit their homework, and grade each other's homework. A screenshot of the homework assessment page of `Blockment` is shown in Figure 2. After all the students in a course upload their solutions to the system, participants in the grading process can click on the "Get an assignment to assess" button to get a randomly selected anonymous homework to grade. The user grades the handout and submits their proposed grade to the system, which will contribute to the student's total grade and be published to the blockchain.

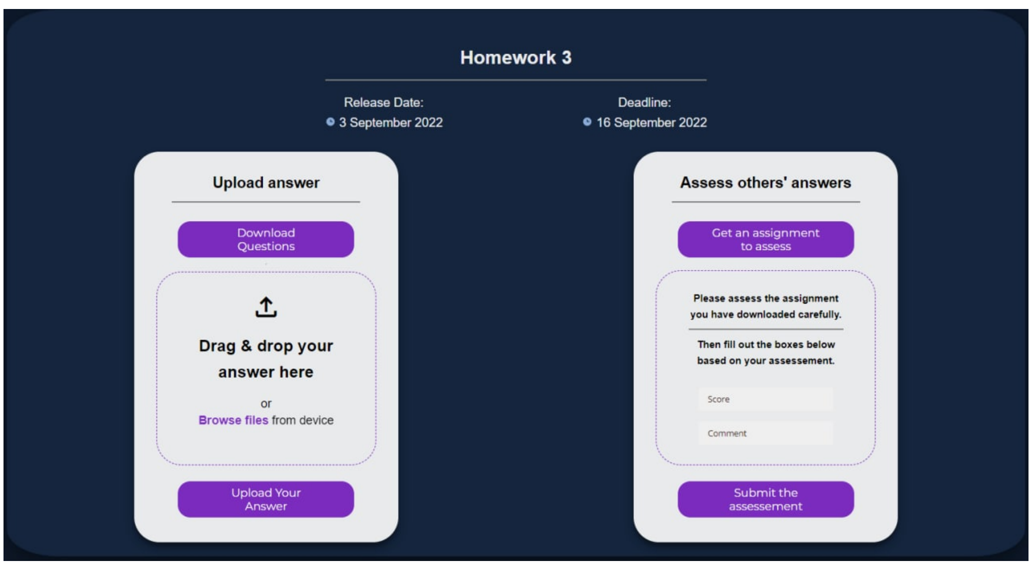

Fig. 2. A screenshot of `Blockment`, showing the page for uploading solutions to homework handouts and assessing handouts of other students enrolled in the same course.

Based on requirement D from Section 4.1, we designed `Blockment` to be a user-centric platform that hides the system's blockchain underpinnings and technical sides. As a result, the students, instructors, and teaching assistants can use `Blockment` without the need to know how to interpret and work with blockchain technologies to use `Blockment`.

[2]https://flask.palletsprojects.com/

The users can interact with `Blockment` as a traditional CW software, and the system will handle the blockchain-related technical tasks in the background.

5 EXPERIMENTAL EVALUATION

In this section, we evaluate our proposed method in an experiment done in classroom environments at a university. By conducting this experiment, we aim to investigate the impact of our method on fairness, reliability, and transparency of the grades, as well as grading motivation and time. To achieve this, we conducted our experiment in two classrooms in the undergraduate computer science program[3] at a West Asian university. We also evaluated the design of `Blockment` qualitatively based on interviews with students. We aimed to conduct our experiments on real students, teaching assistants, and instructors to increase the validity of our experiment design. The participants did not receive compensation; however, they were awarded up to 0.5 additional points (out of 20) in the course after correctly[4] participating in the experiments.

We conducted classroom experiments to test our proposed method on two undergraduate-level classes in a Western Asian university. The first class included 19 students (7 female and 12 male, with an average age of 20.42), and the second class included 33 students (9 female and 24 male, with an average age of 20.03)[5]. Every participant was informed about the purpose of the experiment and that they would be awarded points for their participation via email and in the classroom. We conducted our design evaluation interviews with eleven students (6 male and 5 female, with an average age of 21.29) who all had participated in grading before, either as teaching assistants or as participants in peer grading scenarios.

5.1 Classroom Experiment Design

We designed our classroom study intending to find the effects of our method on five dimensions[6] extracted from previous works (see Section 2.2): 1) reliability of scores, 2) personal bias, 3) grading burden, 4) transparency of grading, and 5) motivation for assessment. As a result, our research questions were as follows:

- **RQ1:** How does our method affect the reliability of the grades?
- **RQ2:** How does our method affect the influence of the personal bias of the assessors on the grades?
- **RQ3:** How does our method affect the grading burden of the graders and the grading speed?
- **RQ4:** How does our method affect the transparency of the grading process?
- **RQ5:** How does our method affect the motivation of the graders?

To answer our research questions, we designed our experiment as follows. First, we presented a summary of our research as a general introduction in the classroom to the students. Then, we asked the students to submit their solutions to the homework problems in their course by uploading files to a Google Drive folder using Google Forms[7]. Students were notified of details on where and how to submit their solutions via emails sent by the course instructors. The handouts were graded in two ways: first, by the instructor in a centralized (traditional) manner, and second, by the

[3]We accept that this sample is not necessarily a true representative of all courses. We specifically note this as a limitation in Section 7.2.
[4]By *correctly*, we mean answering all the questions asked of them in sufficient detail. A researcher checked for answers to see if they included sufficient data or not.
[5]We collected the demographics data via Google Forms.
[6]By which we mean the five extracted limitations from the literature
[7]We collected all the data in Google Drive in order to be able to supervise the process and find investigable results, rather than a fully-anonymous approach, only for the sake of the experiment. Moreover, we considered the student numbers of the students as the indication of their public keys.

students themselves using our method. We adjusted our method to enable supervision of the grading process by the researchers by collecting the grades and a solution in a centralized setting as well, only for the sake of the experiment.

We then compared the grades given by the teaching assistants and the grades given by the students using our method. Each student received their grade at the end of each grading process for that specific student via our system. They could also see the grades of other students at the end. After the student received their score from both the traditional and our new grading methods, we also asked them to fill out a Google Forms questionnaire about their experience with the grading process. The questionnaire link was sent to the students via emails from the course instructors. Besides asking for demographics (age and gender) considering the method described by Hughes et al. [35], we asked four experiment questions (EQs) in the questionnaire. We designed the EQs in a way that we can derive the answers to the RQs from. The EQs are as follows:

- **EQ1:** How fast was the grading process for each of the two methods? (from 0 for very slow to 10 for very fast)
- **EQ2:** How fair and unbiased do you think your grades are for each method? (from 0 for very unfair and biased to 10 for very fair and unbiased)
- **EQ3:** How fair and unbiased do you think other students' grades are for each method?[8] (from 0 for very unfair and biased to 10 for very fair and unbiased)
- **EQ4:** How did you perceive the grading process in our new method? (open question)

All students received the same set of questions. We specifically note that we did not use EQs from previous works and rather designed them on our own to make them fit to our specific method. We choose the EQs such that they can provide us with viable answers to the RQs. We decided that using EQs directly taken from a previous work limited us in coherently representing our solution to the five limitations extracted from the literature. Instead, we provide the EQs in a way that the answers the students give to them can provide us with direct insights into how our method was successful in addressing the five literature-extracted limitations.

5.2 Laboratory Usability Experiment Design

While investigating the effect of our proposed method, our classroom experiment did not evaluate the design and usability of `Blockment`. To investigate the design of `Blockment`, we conducted semi-structured design interviews in the form of laboratory usability tests [19, 20, 56, 84] with eleven students who had participated in peer grading before. All students had the experience of participating in at least three grading tasks in the past. Six students had experienced assessment as a TA duty in other courses, and six were previously enrolled in classes that used traditional peer grading systems[9].

The students were asked to use `Blockment` after a short introduction to the system's goal without formally introducing the blockchain underpinnings. The rules enforced in the system (e.g., the fact that the students can only grade the handouts of other students after they have provided their solutions to the homework) were also not explained explicitly to the students. Two researchers closely observed the student's interactions with the system and took notes about the student's experiences with the system. In the end, the researchers asked the students to answer open questions about their experience with the system and took notes of the answers they provided. Each interview lasted for about 10 to 15 minutes. The usability questions (UQs) asked at the end of each usability interview were as follows:

- **UQ1:** What did you understand about how the assessment method in `Blockment` works?

[8] We also asked questions related to other students to remove any bias in answering questions due to an unexpectedly low or high grade in the homework handouts.
[9] One student had both the TA and peer grading experience.

- **UQ2:** Did you face any difficulty in working with and understanding `Blockment`?
- **UQ3:** What did you especially like about `Blockment`?
- **UQ4:** Did you face any confusion in working with and understanding `Blockment`?
- **UQ5:** What did you dislike about `Blockment`?
- **UQ6:** What do you suggest for improvement in our system?

In designing the UQs, we were inspired by the methods performed in previous qualitative studies evaluating web-based educational tools (e.g. [53, 55]). We modified and adapted the general approach in qualitative studies of web applications to our specific system. We then extracted the UQs ourselves in a way that covers both the positive and the negative points of the system, in addition to asking for improvement suggestions from them.

All participants were asked the same questions.

6 RESULTS

We investigated the effects of our proposed method on the five dimensions of peer assessment (as described in Section 5.1), and the usability of our proposed interface for `Blockment` (as described in Section 5.2). We first present the results of our classroom experiment and then present the results of our laboratory usability experiment.

6.1 Classroom Experiment Results

To answer the research questions in Section 5.1, we analyzed the results of the questionnaire we asked the students to fill out after receiving their grades. We also analyzed the grades given by the instructor traditionally and the grades given by the students using our method. We present the results of our analysis in the following subsections, separated by the research questions.

6.1.1 RQ1: How does our method affect the reliability of the grades? Peer grading has previously shown low reliability of the grades compared to the traditional assessment process [14, 51]. As a result, it is necessary to know if the grades given by our method to students are significantly different compared to the grades given by the traditional method. We used a *Welch Two-Sample T-Test* [49] to find if there is a significant difference between the grades. The mean of the grades given by the instructor with the traditional method was 76.31 out of 100 (SD = 24.06), and the mean of the grades given by the students based on our method was 76.90 out of 100 (SD = 23.41). The p-value obtained from conducting the test was 0.3702 ($p > 0.05$), which means that there is no significant difference between the grades given by the instructor in the traditional way and the grades given by the students using our method. This result doesn't show any evidence that our method is not reliable enough. As the calculated means and STDs are close enough to each other, we claim our method is reliable enough for peer grading. The results from conducting the test are included in Table 1.

6.1.2 RQ2: How does our method affect the influence of the personal bias of the assessors on the grades? Personal bias affects the grades given to the students in conventional peer review settings per the previous works [9]. To see if our method imposes any negative effect on the grades of the students, we ask them to compare the grades obtained from the traditional way with our method and answer the question *How fair and unbiased do you think your grades are for each of the two methods?* with 0 for very unfair and biased, to 10 for very fair and unbiased. The mean of the answers given by the students for the traditional method was 7.88 out of 10 (SD = 1.85), and the mean of the answers given by the students for our method was 8.50 out of 10 (SD = 1.21). The p-value obtained from conducting the *Welch Two-Sample T-Test* [49] was 0.0120 ($p < 0.05$), which means that surprisingly, the measured fairness and unbiasedness of the grades given by the students using our method is significantly higher than the grades given by the instructor in the traditional way.

To remove the effect of the personal bias of the students themselves, we also asked the students to compare the grades of other students between the traditional method and our method and answer the question *How fair and unbiased do you think the grades of other students are for each of the two methods?* with 0 for very unfair and biased, to 10 for very fair and unbiased. The mean of the answers given by the students for the traditional method was 7.38 out of 10 (SD = 2.05), and the mean of the answers given by the students for our method was 8.13 out of 10 (SD = 1.30). The p-value obtained from conducting the *Welch Two-Sample T-Test* [49] was 0.0131 ($p < 0.05$), which reaffirms the claim in the previous comparison. The results thus show that our method is fair and unbiased enough to be used in peer grading. The results from conducting the test are included in Table 1.

6.1.3 RQ3: How does our method affect the grading burden of the graders and the grading speed? According to the existing literature [39], conventional peer grading if supervised directly by moderators instead of a self-reliant network, takes a lot of time. Thus, it is important that our proposed method doesn't take a significantly longer time than the traditional centralized method of grading to make it suitable for use in real classroom environments. We asked the students to compare the time which was spent on grading between receiving the grades from the traditional method and our method by answering the question *How fast was the grading process for each of the two methods, in your opinion?* with 0 for very slow, to 10 for very fast[10]. The mean of the answers given by the students for the traditional method was 5.46 out of 10 (SD = 2.13), and the mean of the answers given by the students for our method was 7.77 out of 10 (SD = 2.34). The p-value obtained from conducting the *Welch Two-Sample T-Test* [49] was 0.000016 ($p < 0.05$), which means that the time spent on grading using the traditional way is significantly higher than the time spent on grading using our new method and our method is less burdensome for the graders. This result shows that our method is not highly time-consuming to be used in assessment settings since it is faster than the conventional centralized method of grading, which is currently being used in many classrooms. The results from conducting the test are included in Table 1.

6.1.4 RQ4: How does our method affect the transparency of the grading process? Previous works have mentioned the low transparency in conventional peer grading methods [62, 73]. As our method aimed to improve and increase transparency, we collected feedback from the students using a form designed in Google Forms platform on how they perceived our method with the open question *How did you perceive the grading process in our new method?*. Two researchers analyzed the answers provided by the students in the same place at the same time, and in case of conflicts in extracting information, they discussed the answers and reached a consensus. The answers given by 46 out of 53 students (86.79%) included feedback on specific aspects of the assessment process (e.g., how the score can be calculated at the end, how the decentralized process helps the grading, potential issues, etc.), which means that the students were able to understand the grading process, and thus, the transparency of the grading process was increased. This result shows that our method is transparent enough to be used in peer grading.

6.1.5 RQ5: How does our method affect the motivation of the graders? Previous researchers have demonstrated the necessity for external motivation to get students involved in peer grading [85]. The instructors in the two courses on which we conducted our experiment agreed to provide bonus points to students successfully participating in the grading process and successfully completing the questionnaire. From the 1.0 bonus point designated for each student, 0.2 was

[10]We aimed to measure the grading speed qualitatively, since a) we didn't have precise data for the time it took for the instructor to grade each handout separately and classified according to the name of the respective student, and b) we wanted to measure what feeling did the students have on the speed of grading and if they feel the shorter grading time has benefited them in any specific way or not.

Measure	Mean Traditional	Mean Our Method	SD Traditional	SD Our Method	p-value
Grades (from 100)	76.31	76.90	24.06	23.41	0.3702
Fairness of student's grades (from 10)	7.88	8.50	1.85	1.21	0.0120 *
Fairness of others' grades (from 10)	7.38	8.13	2.05	1.30	0.0131 *
Speed of grading (from 10)	5.46	7.77	2.13	2.34	0.000016 ***

Table 1. Mean, standard deviation, and the two-tailed p-value of conducting *Welch Two-Sample T-Tests* on various measures from the survey results. ***p < 0.001, **p < 0.01, *p < 0.05

awarded to all participating students equally, and the rest (0.8) was awarded proportionally to the grades they received from our peer grading method. This motivation is a simulation of our method in the experiment; in the real method, the results obtained from the peer graders also have an effect on the total score of the students, contributing to their motivation for participating in the assessment process. Our results also prove the motivation was sufficient in order to persuade students to provide detailed feedback and responses; 51 out of 53 students (96.23%) provided comprehensive and detailed answers to the open question in the questionnaire.

6.2 Lab Experiment Results

To find insights on the usability of `Blockment` from the viewpoint of students, we tracked the interactions of eleven students (S_1 to S_{11}) with `Blockment` and asked them the six questions indicated in Section 5.2 in our laboratory experiment. The researchers took note of all the interactions as well as the oral answers to the questions asked by the participants. The participants were advised to "think aloud" [11] while working with the system so that they enable researchers to track better the interactions and how the students perceive the user flow of `Blockment`. After the interviews, two researchers annotated the collected notes at the same time and resolved any conflicts as they were brought up. The annotations consisted of the labels *positive point*, *negative points*, *confusing aspects*, and *suggestions*. For each label, the sample comments were collected and summarized to find the main insights on the design of the `Blockment`. Generally, all the participants liked the design of `Blockment` and were able to become familiar quickly with how they should work with the system.

6.2.1 Positive Points. Eight participants (S_2 to S_9) mentioned a resemblance of the general user experience flow to the currently-in-use CW system of their university. Two participants (S_1 and S_{10}) mentioned the familiarity of the interface of `Blockment` and its similarity to existing MOOC platforms. Four participants (S_1, S_2, S_5, and S_{11}) mentioned the clean and intuitive interface of `Blockment`, and three participants (S_1, S_5, and S_7) specifically mentioned the benefit of allocating a separate, all-in-one page for every homework, in which they can both submit their own solutions and grade those of others, in providing an intuitive interface.

6.2.2 Negative Points. Two participants (S_3 and S_4) expected a search functionality in courses and homework titles, which was not existent in the original design. One participant (S_6) didn't like the system's flow, found it unnecessarily complicated, and preferred to upload the files from the main page without needing to navigate to the inner pages. Five participants (S_5 to S_9) expected more detailed descriptions of the assessment process to increase transparency by means of the user interface.

6.2.3 Confusing Aspects. Two participants (S_1 and S_5) were initially confused by the assessment process in `Blockment` and could not discover that they first needed to submit a solution themselves before starting to grade the others. One participant (S_{11}) could not find how our method's assessment process works and differs from the traditional peer grading processes by looking at the system alone, and needed more explanation from the researchers to find out how the system works.

6.2.4 Suggestions. Three participants (S_1, S_7, and S_{10}) mentioned the need for a system-wide help functionality, which is evident in many other systems [94], but didn't exist in `Blockment`. One participant (S_4) suggested adding a learner's dashboard where users can see which courses they have submitted their handouts for, which courses they have to grade handouts for, and how much they have progressed in their enrolled courses. Two participants (S_3 and S_6) suggested embedding `Blockment` as an extension to existing CW systems (e.g., Moodle[11]) as compared to a separate system, to maximize the familiarity with the system when using our new assessment method.

7 DISCUSSION

In this research, we explore the previous works and available literature on computer-supported peer assessment. We find five of the limitations considered in previous works, namely *low reliability of scores*, *lack of transparency*, *effect of personal bias*, *high grading burden*, and *lack of strong motivation*. To find a viable solution to these limitations, the current study investigates the benefits and challenges of using a peer grading system based on blockchain technology among students and in educational environments. As our method is new, there is also the need to investigate how this method can be embedded in a user-centric interface. Thus, in addition to providing and evaluating our peer assessment method, we design and test `Blockment` and provide details on how to embed our method in a user-centric tool.

Our evaluation of the new method proposed by us indicates that our method has had positive benefits on various measurable dimensions of grading procedures, including grade reliability, the bias of assessors, grading speed, transparency of the assessment process, and motivation of the graders. Moreover, our laboratory experiments show that the design of `Blockment` is generally intuitive to the students and that the students can quickly become familiar with our system. Comparing the grades in the two methods (our new method and the traditional centralized method) indicates that our methods did not pose any significant inflation in the grades. Also, students claimed the grades in the new method were significantly fairer than the traditional way and received their grades significantly faster than the previous approach. They were also motivated to participate in the grading process using our approach to peer grading and demonstrated a high understanding of how the grading process works in our method.

In summary, the results from our experiment indicate:

- There is no significant difference between the grades given to the students in our method versus the traditional method, indicating that our method doesn't affect the grade **reliability**.
- There is a significantly higher measured fairness and unbiasedness in the grades in our method compared to the traditional method, indicating that our method is **fair** and **unbiased** in the peer grading context.

[11]https://moodle.org/

- The traditional way takes a significantly longer time for the grading process, indicating that our method is a **faster** approach to grading.
- The answers to the open-ended questions asked from the students specifically includes feedback signifying a deep understanding of the grading process, indicating that our method is considered **transparent** in a peer grading setting.
- Students provided comprehensive answers and were motivated in participating in the experiment due to the bonus score approach we took into account in the method. Consequently, the **motivation** was considered sufficient although the students were not compensated with money for participating in the grading process.
- The results of the lab experiment for the design of `Blockment` indicate an overall beneficial **user experience** flow, as well as providing suggestions on how to further improve systems used for peer grading contexts.

7.1 Theoretical and Practical Contributions

Former research studies have investigated systems whose major theories are based on methods to make assessment fairer with peer grading. However, this study tried to introduce a modern method to design a system that aims to make assessment fairer, more transparent, faster, more instructive for students, and more reliable. We showed in our experiments that the results of the questionnaires collected from the students proved that they believe that our method is fairer than the conventional grading method. Moreover, as the main attributes of a blockchain-based system are transparency, our method was also considered transparent by the students. Not only the proposed method is fairer, but also it is faster. Finally, our suggested system is more reliable, and the process of assessing the papers could be trusted in a course based on the experience of the classes on which we tested our method.

Therefore, our research makes several contributions to the literature on HCI and CSCL. First, we propose a new peer grading method based on blockchain technology, which, according to our study, is more reliable, more transparent, fairer, faster, and more motivating than the traditional centralized approaches. Second, we identify the potential drawbacks of adopting a decentralized system and make accommodations for them in our blockchain backbone architecture. Third, we design and implement `Blockment`, a user-centric tool for peer grading, and evaluate it in a laboratory experiment. We also provide a collection of user feedback on the design of `Blockment`, which can be used as a guideline for future user-centric peer grading systems.

7.2 Limitations and Future Work

Nevertheless, there are known limitation in our research work, especially in how we measure the metrics above. For example, we only compared the grades in our method to traditional centralized approaches and did not perform any comparison with the other conventional peer grading methods. While the comparison with centralized settings demonstrates that our method does not have any specific adverse effect on the grades, future works can compare with more methods to further investigate our proposed method in educational settings. In addition, we measured fairness using two questions (asking for the fairness of the student's grades, as well as the grades of the classmates, for each student) to reduce the personal bias of the students; however, more study should be done to find if this is the premier approach to reduce the personal bias of the participants. Moreover, in our method, we trust the students not to write their names on their solutions, and we checked the solutions manually to make sure they didn't violate this rule. However, we call for researchers to find how to enforce this rule in real-world large-scale educational settings.

Also, as we didn't have access to precise data on how long the instructor took to grade each handout separately (classified according to the respective student's name), we aimed to measure the grading speed qualitatively. Future studies can investigate the effect of our method on the grading speed in more details and/or quantitatively. Moreover, we did not ask any *specific* questions in the survey regarding the motivation of the students participating in our study, so our claims on the student's motivation lack quantitative backing data. We suggest future researchers add specific queries on the participants' motivation to participate in our peer grading process.

Another limitation is the approach to detect fraud in the system. The students may decide to misjudge the students' papers, and all the students of the class score others may collaboratively a certain grade; therefore, in that case, the scores would not be correct. To mitigate this issue, we propose that the instructor and the teaching assistants participate in the grading process as well, and if the difference between their scores and those of the students is higher than a certain threshold, it would be detected as fraud. Moreover, if the number of students in the class tends to be large, the possibility of this kind of fraud being occurred will become less. We invite future researchers to find specific rules to implement in such a system, so that fraud in the grading process is detected before being conducted and submitted to the blockchain. Finally, it is noteworthy to mention that our study has been done in a limited scale, for example, including only a limited number of classrooms and students in the undergraduate program of our university; thus, more detailed and in-depth investigations are needed to sufficiently prove the claimed contributions.

8 CONCLUSION

In this paper, we proposed a new peer grading method based on blockchain technology and the previous literature on HCI and CSCL by addressing the principal shortcomings of conventional peer grading approaches, as noted in the previous works. In addition, we extracted specific principles and design concepts for a user-centric design incorporating our method and developed `Blockment` based on the collected ideas. We tested our method in a classroom experiment, which showed high reliability of grades, low personal bias of assessors, high speed of grading, high transparency of the assessment process, and high motivation of participants. We also conducted a laboratory experiment to evaluate the design of `Blockment` and collected and clustered user feedback on the design of our system. The results of our experiments show our method and our implemented system as up-and-coming viable solutions for peer assessment processes in educational environments.

REFERENCES

[1] Nurzhan Zhumabekuly Aitzhan and Davor Svetinovic. 2016. Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams. *IEEE Transactions on Dependable and Secure Computing* 15, 5 (2016), 840–852.

[2] Ali Alammary, Samah Alhazmi, Marwah Almasri, and Saira Gillani. 2019. Blockchain-based applications in education: A systematic review. *Applied Sciences* 9, 12 (2019), 2400.

[3] Mustafa Cem Aldag. 2020. The use of blockchain technology in agriculture. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie/Cracow Review of Economics and Management* 4 (982) (2020), 7–17.

[4] Mohammed AlShamsi, Said A Salloum, Muhammad Alshurideh, and Sherief Abdallah. 2021. Artificial intelligence and blockchain for transparency in governance. In *Artificial intelligence for sustainable development: Theory, practice and future applications*. Springer, 219–230.

[5] Rodelio Arenas and Proceso Fernandez. 2018. CredenceLedger: a permissioned blockchain for verifiable academic credentials. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. IEEE, 1–6.

[6] Rawia Bdiwi, Cyril De Runz, Sami Faiz, and Arab Ali Cherif. 2018. A blockchain based decentralized platform for ubiquitous learning environment. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 90–92.

[7] Nelson Bore, Samuel Karumba, Juliet Mutahi, Shelby Solomon Darnell, Charity Wayua, and Komminist Weldemariam. 2017. Towards blockchain-enabled school information hub. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*. 1–4.

[8] Steven Bradley. 2019. Addressing Bias to Improve Reliability in Peer Review of Programming Coursework. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '19)*. Association for Computing Machinery, New York, NY, USA, Article 19, 10 pages. https://doi.org/10.1145/3364510.3364523

[9] Clare Brindley and Susan Scoffield. 1998. Peer Assessment in Undergraduate Programmes. *Teaching in Higher Education* 3, 1 (1998), 79–90. https://doi.org/10.1080/1356215980030106 arXiv:https://doi.org/10.1080/1356215980030106

[10] Hou Pong Chan and Irwin King. 2017. Leveraging Social Connections to Improve Peer Assessment in MOOCs. In *Proceedings of the 26th International Conference on World Wide Web Companion* (Perth, Australia) *(WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 341–349. https://doi.org/10.1145/3041021.3054165

[11] Elizabeth Charters. 2003. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal* 12, 2 (2003).

[12] Jingchun Chen. 2021. The Effectiveness of Peer Assessment in EFL Blended Learning Environments. In *2021 2nd International Conference on Computers, Information Processing and Advanced Education* (Ottawa, ON, Canada) *(CIPAE 2021)*. Association for Computing Machinery, New York, NY, USA, 566–568. https://doi.org/10.1145/3456887.3457015

[13] Winnie Cheng and Martin Warren. 1997. Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education* 22, 2 (1997), 233–239. https://doi.org/10.1080/03075079712331381064 arXiv:https://doi.org/10.1080/03075079712331381064

[14] Winnie Cheng and Martin Warren. 1999. Peer and Teacher Assessment of the Oral and Written Tasks of a Group Project. *Assessment & Evaluation in Higher Education* 24, 3 (1999), 301–314. https://doi.org/10.1080/0260293990240304 arXiv:https://doi.org/10.1080/0260293990240304

[15] Avi J. Cohen and Andrea L. Williams. 2019. Scalable, scaffolded writing assignments with online peer review in a large introductory economics course. *The Journal of Economic Education* 50, 4 (2019), 371–387. https://doi.org/10.1080/00220485.2019.1654951 arXiv:https://doi.org/10.1080/00220485.2019.1654951

[16] Michael Crosby, Pradan Pattanayak, Sanjeev Verma, Vignesh Kalyanaraman, et al. 2016. Blockchain technology: Beyond bitcoin. *Applied Innovation* 2, 6-10 (2016), 71.

[17] Phil Davies. 2000. Computerized Peer Assessment. *Innovations in Education and Training International* 37, 4 (2000), 346–355. https://doi.org/10.1080/135580000750052955 arXiv:https://doi.org/10.1080/135580000750052955

[18] Tamara Denning, Michael Kelly, David Lindquist, Roshni Malani, William G. Griswold, and Beth Simon. 2007. Lightweight Preliminary Peer Review: Does in-Class Peer Review Make Sense?. In *Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education* (Covington, Kentucky, USA) *(SIGCSE '07)*. Association for Computing Machinery, New York, NY, USA, 266–270. https://doi.org/10.1145/1227310.1227406

[19] Joseph S Dumas, Joseph S Dumas, and Janice Redish. 1999. *A practical guide to usability testing*. Intellect books.

[20] Joseph S Dumas and Jean E Fox. 2007. Usability testing: Current practice and future directions. In *The human-computer interaction handbook*. CRC Press, 1155–1176.

[21] Chris Elsden, Arthi Manohar, Jo Briggs, Mike Harding, Chris Speed, and John Vines. 2018. Making Sense of Blockchain Applications: A Typology for HCI. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174032

[22] Stephen Fallows and Balasubramanyan Chandramohan. 2001. Multiple Approaches to Assessment: Reflections on use of tutor, peer and self-assessment. *Teaching in Higher Education* 6, 2 (2001), 229–246. https://doi.org/10.1080/13562510120045212 arXiv:https://doi.org/10.1080/13562510120045212

[23] George Filippakis, Katerina Georgouli, and Cleo Sgouropoulou. 2015. The Design of a Web-Based System to Support the Distribution of Assignments for Distance Peer Evaluation. In *Proceedings of the 19th Panhellenic Conference on Informatics* (Athens, Greece) *(PCI '15)*. Association for Computing Machinery, New York, NY, USA, 400–405. https://doi.org/10.1145/2801948.2802014

[24] Edward F. Gehringer. 2001. Electronic Peer Review and Peer Grading in Computer-Science Courses. In *Proceedings of the Thirty-Second SIGCSE Technical Symposium on Computer Science Education* (Charlotte, North Carolina, USA) *(SIGCSE '01)*. Association for Computing Machinery, New York, NY, USA, 139–143. https://doi.org/10.1145/364447.364564

[25] Edward F. Gehringer, Donald D. Chinn, Manuel A. Pérez-Quiñones, and Mark A. Ardis. 2005. Using Peer Review in Teaching Computing. In *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education* (St. Louis, Missouri, USA) *(SIGCSE '05)*. Association for Computing Machinery, New York, NY, USA, 321–322. https://doi.org/10.1145/1047344.1047455

[26] Katerina Georgouli, C. Sgouropoulou, Ilias Skalkidis, and Charilaos Tsetsekas. 2012. Introducing a Collaborative Peer-Evaluation Learning Model in Higher Education Programming-Based Courses. *Proceedings of the 2012 16th Panhellenic Conference on Informatics, PCI 2012*, 399–404. https://doi.org/10.1109/PCi.2012.72

[27] Hanna Halaburda. 2018. Blockchain Revolution without the Blockchain? *Commun. ACM* 61, 7 (jun 2018), 27–29. https://doi.org/10.1145/3225619

[28] John Hamer, Catherine Kell, and Fiona Spence. 2007. Peer Assessment Using Aropä. In *Proceedings of the Ninth Australasian Conference on Computing Education - Volume 66* (Ballarat, Victoria, Australia) *(ACE '07)*. Australian Computer Society, Inc., AUS, 43–54.

[29] Meng Han, Zhigang Li, Jing He, Dalei Wu, Ying Xie, and Asif Baba. 2018. A novel blockchain-based education records verification solution. In *Proceedings of the 19th annual SIG conference on information technology education*. 178–183.

[30] Yu He, Xinying Hu, and Guangzhong Sun. 2019. A Cognitive Diagnosis Framework Based on Peer Assessment. In *Proceedings of the ACM Turing Celebration Conference - China* (Chengdu, China) *(ACM TURC '19)*. Association for Computing Machinery, New York, NY, USA, Article 78, 6 pages. https://doi.org/10.1145/3321408.3322850

[31] Thomas Hepp, Alexander Schoenhals, Christopher Gondek, and Bela Gipp. 2018. OriginStamp: A blockchain-backed system for decentralized trusted timestamping. *it-Information Technology* 60, 5-6 (2018), 273–281.

[32] Janette R. Hill, Liyan Song, and Richard E. West. 2009. Social Learning Theory and Web-Based Learning Environments: A Review of Research and Discussion of Implications. *American Journal of Distance Education* 23, 2 (2009), 88–103. https://doi.org/10.1080/08923640902857713 arXiv:https://doi.org/10.1080/08923640902857713

[33] Marko Hölbl, Aida Kamisalić, Muhamed Turkanović, Marko Kompara, Blaž Podgorelec, and Marjan Heričko. 2018. EduCTX: an ecosystem for managing digital micro-credentials. In *2018 28th EAEEIE Annual Conference (EAEEIE)*. IEEE, 1–9.

[34] Tani Hossain, Tasniah Mohiuddin, AM Hasan, Muhammad Nazrul Islam, and Syed Akhter Hossain. 2020. Designing and developing graphical user interface for the MultiChain blockchain: towards incorporating HCI in blockchain. In *International Conference on Intelligent Systems Design and Applications*. Springer, 446–456.

[35] Jennifer L Hughes, Abigail A Camden, Tenzin Yangchen, et al. 2016. Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research* 21, 3 (2016), 138–151.

[36] Suhan Jiang and Jie Wu. 2022. A reward response game in the blockchain-powered federated learning system. *International Journal of Parallel, Emergent and Distributed Systems* 37, 1 (2022), 68–90.

[37] Bojana Koteska, Elena Karafiloski, and Anastas Mishev. 2017. Blockchain implementation quality challenges: a literature. In *SQAMIA 2017: 6th workshop of software quality, analysis, monitoring, improvement, and applications*, Vol. 1938. 8–8.

[38] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer. 2013. Peer and Self Assessment in Massive Online Classes. *ACM Trans. Comput.-Hum. Interact.* 20, 6, Article 33 (dec 2013), 31 pages. https://doi.org/10.1145/2505057

[39] Chinmay E. Kulkarni, Richard Socher, Michael S. Bernstein, and Scott R. Klemmer. 2014. Scaling Short-Answer Grading by Combining Peer Assessment with Algorithmic Scoring. In *Proceedings of the First ACM Conference on Learning @ Scale Conference* (Atlanta, Georgia, USA) *(L@S '14)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/2556325.2566238

[40] Yi-Chieh Lee and Wai-Tat Fu. 2019. Supporting Peer Assessment in Education with Conversational Agents. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 7–8. https://doi.org/10.1145/3308557.3308695

[41] Erno Lehtinen, Kai Hakkarainen, Lasse Lipponen, Marjaana Rahikainen, and Hanni Muukkonen. 1999. Computer supported collaborative learning: A review. *The JHGI Giesbers reports on education* 10 (1999), 1999.

[42] Jinrong Li and Mimi Li. 2018. Turnitin and peer review in ESL academic writing classrooms. (2018).

[43] Sunny San-Ju Lin, E. Z.-F. Liu, and Shyan-Ming Yuan. 2001. Web Based Peer Assessment: Attitude and Achievement. *IEEE Trans. on Educ.* 44, 2 (may 2001), 13 pp. https://doi.org/10.1109/13.925865

[44] Yu-Pin Lin, Joy R Petway, Wan-Yu Lien, and Josef Settele. 2018. Blockchain with artificial intelligence to efficiently manage water use under climate change. , 34 pages.

[45] Eric Zhi-Feng Liu, S. S.J. Lin, Chi-Huang Chiu, and Shyan-Ming Yuan. 2001. Web-Based Peer Review: The Learner as Both Adapter and Reviewer. *IEEE Trans. on Educ.* 44, 3 (aug 2001), 246–251. https://doi.org/10.1109/13.940995

[46] Qin Liu, Qingchen Guan, Xiaowen Yang, Hongming Zhu, Gill Green, and Shaohan Yin. 2018. Education-industry cooperative system based on blockchain. In *2018 1st IEEE international conference on hot information-centric networking (HotICN)*. IEEE, 207–211.

[47] David Lizcano, Juan A Lara, Bebo White, and Shadi Aljawarneh. 2020. Blockchain-based approach to create a model of trust in open and ubiquitous higher education. *Journal of Computing in Higher Education* 32, 1 (2020), 109–134.

[48] Yanxin Lu, Joe Warren, Christopher Jermaine, Swarat Chaudhuri, and Scott Rixner. 2015. Grading the Graders: Motivating Peer Graders in a MOOC. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 680–690. https://doi.org/10.1145/2736277.2741649

[49] Zhenqiu Lu and Ke-Hai Yuan. 2010. Welch's t test. *Encyclopedia of research design* (2010), 1620–1623.

[50] Heng Luo, Anthony Robinson, and Jae-Young Park. 2014. Peer grading in a MOOC: Reliability, validity, and perceived effects. *Online Learning Journal* 18, 2 (2014).

[51] Douglas Magin and Phil Helmore. 2001. Peer and Teacher Assessments of Oral Presentation Skills: How reliable are they? *Studies in Higher Education* 26, 3 (2001), 287–298. https://doi.org/10.1080/03075070120076264 arXiv:https://doi.org/10.1080/03075070120076264

[52] Roberto Mavilia and Roberta Pisani. 2022. Blockchain for agricultural sector: The case of South Africa. *African Journal of Science, Technology, Innovation and Development* 14, 3 (2022), 845–851.

[53] Paola Mejia-Domenzain, Eva Laini, Seyed Parsa Neshaei, Thiemo Wambsganss, and Tanja Käser. 2023. Visualizing Self-Regulated Learner Profiles in Dashboards: Design Insights from Teachers. *arXiv preprint arXiv:2305.16851* (2023).

[54] Alexander Mikroyannidis, John Domingue, Michelle Bachler, and Kevin Quick. 2018. A learner-centred approach for lifelong learning powered by the blockchain. In *EdMedia+ innovate learning*. Association for the Advancement of Computing in Education (AACE), 1388–1393.

[55] Martijn Millecamp, Francisco Gutiérrez, Sven Charleer, Katrien Verbert, and Tinne De Laet. 2018. A qualitative evaluation of a learning dashboard to support advisor-student dialogues. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 56–60.

[56] Rolf Molich, Ann Damgaard Thomsen, Barbara Karyukina, Lars Schmidt, Meghan Ede, Wilma van Oel, and Meeta Arcuri. 1999. Comparative evaluation of usability tests. In *CHI'99 extended abstracts on Human factors in computing systems*. 83–84.

[57] 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. https://doi.org/10.1177/1469787414527391 arXiv:https://doi.org/10.1177/1469787414527391

[58] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review* (2008), 21260.

[59] Tricia J. Ngoon, Rachel Chen, Amit Deutsch, and Sean Lip. 2016. Oppia: A Community of Peer Learners to Make Conversational Learning Experiences. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (San Francisco, California, USA) *(CSCW '16 Companion)*. Association for Computing Machinery, New York, NY, USA, 73–76. https://doi.org/10.1145/2818052.2874328

[60] Shunsuke Noguchi and Naomi Fujimura. 2015. Implementation and Experience of the Online Peer Grading System for Our Real Class. In *Proceedings of the 2015 ACM SIGUCCS Annual Conference* (St. Petersburg, Florida, USA) *(SIGUCCS '15)*. Association for Computing Machinery, New York, NY, USA, 125–128. https://doi.org/10.1145/2815546.2815581

[61] Joonsuk Park and Kimberley Williams. 2016. The Effects of Peer- and Self-Assessment on the Assessors. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (Memphis, Tennessee, USA) *(SIGCSE '16)*. Association for Computing Machinery, New York, NY, USA, 249–254. https://doi.org/10.1145/2839509.2844602

[62] Meghan Perdue and Jessica Sandland. 2022. Evaluating the impact of transparency on peer review quality in MOOCs. In *2022 IEEE Learning with MOOCS (LWMOOCS)*. 101–107. https://doi.org/10.1109/LWMOOCS53067.2022.9928020

[63] Chris Phielix, Frans J. Prins, and Paul A. Kirschner. 2010. Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior* 26, 2 (2010), 151–161. https://doi.org/10.1016/j.chb.2009.10.011

[64] Chris Piech, Jonathan Huang, Zhenghao Chen, Chuong Do, Andrew Ng, and Daphne Koller. 2013. Tuned models of peer assessment in MOOCs. *arXiv preprint arXiv:1307.2579* (2013).

[65] Aleksandra Prikhodko and Elvir Akhmetshin. 2022. Peer-to-Peer Grading as a Way to Improve Educational Process Efficiency for Certain Industrial Sectors under Digitalization of Education at Universities. In *IV International Scientific and Practical Conference* (St.Petersburg, Russian Federation) *(DEFIN-2021)*. Association for Computing Machinery, New York, NY, USA, Article 22, 5 pages. https://doi.org/10.1145/3487757.3490860

[66] Frans J. Prins, Dominique M. A. Sluijsmans, Paul A. Kirschner, and Jan-Willem Strijbos. 2005. Formative peer assessment in a CSCL environment: a case study. *Assessment & Evaluation in Higher Education* 30, 4 (2005), 417–444. https://doi.org/10.1080/02602930500099219 arXiv:https://doi.org/10.1080/02602930500099219

[67] Ricardo Raimundo and Albérico Rosário. 2021. Blockchain system in the higher education. *European Journal of Investigation in Health, Psychology and Education* 11, 1 (2021), 276–293.

[68] Ken Reily, Pam Ludford Finnerty, and Loren Terveen. 2009. Two Peers Are Better than One: Aggregating Peer Reviews for Computing Assignments is Surprisingly Accurate. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '09)*. Association for Computing Machinery, New York, NY, USA, 115–124. https://doi.org/10.1145/1531674.1531692

[69] Jennifer M Robinson. 2002. In search of fairness: An application of multi-reviewer anonymous peer review in a large class. *Journal of Further and Higher Education* 26, 2 (2002), 183–192.

[70] Philip M Sadler and Eddie Good. 2006. The impact of self-and peer-grading on student learning. *Educational assessment* 11, 1 (2006), 1–31.

[71] Mike Sharples and John Domingue. 2016. The blockchain and kudos: A distributed system for educational record, reputation and reward. In *European conference on technology enhanced learning*. Springer, 490–496.

[72] Jirarat Sitthiworachart and Mike Joy. 2004. Effective Peer Assessment for Learning Computer Programming. In *Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education* (Leeds, United Kingdom) *(ITiCSE '04)*. Association for Computing Machinery, New York, NY, USA, 122–126. https://doi.org/10.1145/1007996.1008030

[73] Holly Smith, Ali Cooper, and Les Lancaster. 2002. Improving the Quality of Undergraduate Peer Assessment: A Case for Student and Staff Development. *Innovations in Education and Teaching International* 39, 1 (2002), 71–81. https://doi.org/10.1080/13558000110102904 arXiv:https://doi.org/10.1080/13558000110102904

[74] Harald Sondergaard. 2009. Learning from and with Peers: The Different Roles of Student Peer Reviewing. In *Proceedings of the 14th Annual ACM SIGCSE Conference on Innovation and Technology in Computer Science Education* (Paris, France) *(ITiCSE '09)*. Association for Computing Machinery, New York, NY, USA, 31–35. https://doi.org/10.1145/1562877.1562893

[75] Remya Stephen and Aneena Alex. 2018. A review on blockchain security. In *IOP Conference Series: Materials Science and Engineering*, Vol. 396. IOP Publishing, 012030.

[76] Brent Strong, Mark Davis, and Val Hawks. 2004. Self-Grading In Large General Education Classes: A Case Study. *College Teaching* 52, 2 (2004), 52–57. https://doi.org/10.3200/CTCH.52.2.52-57 arXiv:https://doi.org/10.3200/CTCH.52.2.52-57

[77] Xiaotian Su, Thiemo Wambsganss, Roman Rietsche, Seyed Parsa Neshaei, and Tanja Käser. 2023. Reviewriter: AI-Generated Instructions For Peer Review Writing. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*. Association for Computational Linguistics, Toronto, Canada, 57–71. https://aclanthology.org/2023.bea-1.5

[78] Sarah L. Sullivan. 1994. Reciprocal Peer Reviews. In *Proceedings of the Twenty-Fifth SIGCSE Symposium on Computer Science Education* (Phoenix, Arizona, USA) *(SIGCSE '94)*. Association for Computing Machinery, New York, NY, USA, 314–318. https://doi.org/10.1145/191029.191158

[79] Huang-Chih Sung. 2018. When open source software encounters patents: blockchain as an example to explore the dilemma and solutions. *J. Marshall Rev. Intell. Prop. L.* 18 (2018), v.

[80] Atima Tharatipyakul and Suporn Pongnumkul. 2021. User Interface of Blockchain-Based Agri-Food Traceability Applications: A Review. *IEEE Access* 9 (2021), 82909–82929. https://doi.org/10.1109/ACCESS.2021.3085982

[81] Daniel Toll and Anna Wingkvist. 2017. How Tool Support and Peer Scoring Improved Our Students' Attitudes Toward Peer Reviews. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education* (Bologna, Italy) *(ITiCSE '17)*. Association for Computing Machinery, New York, NY, USA, 311–316. https://doi.org/10.1145/3059009.3059059

[82] Keith Topping. 1998. Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research* 68, 3 (1998), 249–276. https://doi.org/10.3102/00346543068003249 arXiv:https://doi.org/10.3102/00346543068003249

[83] A. Trivedi, Dulal C. Kar, and Holly Patterson-McNeill. 2003. Automatic Assignment Management and Peer Evaluation. *J. Comput. Sci. Coll.* 18, 4 (apr 2003), 30–37.

[84] Tom Tullis, Stan Fleischman, Michelle McNulty, Carrie Cianchette, and Margaret Bergel. 2002. An empirical comparison of lab and remote usability testing of web sites. In *Usability Professionals Association Conference*.

[85] Scott Turner, Manuel A. Pérez-Quiñones, Stephen Edwards, and Joseph Chase. 2011. Student Attitudes and Motivation for Peer Review in CS2. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education* (Dallas, TX, USA) *(SIGCSE '11)*. Association for Computing Machinery, New York, NY, USA, 347–352. https://doi.org/10.1145/1953163.1953268

[86] Sarah Underwood. 2016. Blockchain beyond bitcoin. *Commun. ACM* 59, 11 (2016), 15–17.

[87] I. Van den Berg, Wilfried Admiraal, and Albert Pilot. 2006. Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education* 31 (06 2006), 341–356. https://doi.org/10.1080/03075070600680836

[88] P Vinothiyalakshmi, C Muralidharan, Y Mohamed Sirajudeen, and R Anitha. 2022. Digitized Land Registration Using Blockchain Technology. In *Blockchain Technology*. CRC Press, 73–86.

[89] Tim Vogelsang and Lara Ruppertz. 2015. On the Validity of Peer Grading and a Cloud Teaching Assistant System. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (Poughkeepsie, New York) *(LAK '15)*. Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/2723576.2723633

[90] Huaqun Wang, Debiao He, Zhe Liu, and Rui Guo. 2019. Blockchain-based anonymous reporting scheme with anonymous rewarding. *IEEE Transactions on Engineering Management* 67, 4 (2019), 1514–1524.

[91] Bin Wu and Yinsheng Li. 2018. Design of evaluation system for digital education operational skill competition based on blockchain. In *2018 IEEE 15th international conference on e-business engineering (ICEBE)*. IEEE, 102–109.

[92] Junfeng Xie, F. Richard Yu, Tao Huang, Renchao Xie, Jiang Liu, and Yunjie Liu. 2019. A Survey on the Scalability of Blockchain Systems. *IEEE Network* 33, 5 (2019), 166–173. https://doi.org/10.1109/MNET.001.1800290

[93] Nur Syafiqah Yaccob, Melor Md Yunus, and Harwati Hashim. 2022. The Integration of Global Competence Into Malaysian English as a Second Language Lessons for Quality Education (Fourth United Nations Sustainable Development Goal). *Frontiers in Psychology* 13 (2022).

[94] Su-Fang Yeh, Meng-Hsin Wu, Tze-Yu Chen, Yen-Chun Lin, XiJing Chang, You-Hsuan Chiang, and Yung-Ju Chang. 2022. How to Guide Task-Oriented Chatbot Users, and When: A Mixed-Methods Study of Combinations of Chatbot Guidance Types and Timings. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 488, 16 pages. https://doi.org/10.1145/3491102.3501941

[95] Willie Yip. 2004. Web-Based Support for Peer Tutoring. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT '04)*. IEEE Computer Society, USA, 619–623.

[96] Ahmed Mohamed Fahmy Yousef, Usman Wahid, Mohamed Amine Chatti, Ulrik Schroeder, and Marold Wosnitza. 2015. The Effect of Peer Assessment Rubrics on Learners' Satisfaction and Performance Within a Blended MOOC Environment. In *Proceedings of the 7th International Conference on Computer Supported Education - Volume 2* (Lisbon, Portugal) *(CSEDU 2015)*. SCITEPRESS - Science and Technology Publications, Lda, Setubal, PRT, 148–159. https://doi.org/10.5220/0005495501480159

[97] Javad Zarrin, Hao Wen Phang, Lakshmi Babu Saheer, and Bahram Zarrin. 2021. Blockchain for decentralization of internet: prospects, trends, and challenges. *Cluster Computing* 24, 4 (2021), 2841–2866.

[98] Li Zhang and Jianbo Xu. 2022. Blockchain-based anonymous authentication for traffic reporting in VANETs. *Connection Science* 34, 1 (2022), 1038–1065.

[99] Jiemin Zhong, Haoran Xie, Di Zou, and Dickson KW Chui. 2018. A blockchain model for word-learning systems. In *2018 5th international conference on behavioral, economic, and socio-cultural computing (BESC)*. IEEE, 130–131.