DSI DATA SCIENCE AND INFORMETRICS
VOL.1, NO.4, NOV. 2021

Toward a data analytics–based approach for computing and predicting context–sensitive teaching effectiveness

Kun Tian[a], Ying He[b*], Xiaoran Xu[b], Jiangyang Fu[c]

a. College of Software, Tianjin Normal University, Tianjin, China
b. School of Management, Tianjin Normal University, Tianjin, China
c. School of Information Management, Nanjing University, Nanjing, Jiangsu Province, China

ABSTRACT

The traditional student-oriented course evaluation has been the major assessment method on teaching effectiveness worldwide. Useful as it is, it has been widely and continuously criticized for not being a fair, accurate, and reliable measurement. In search of a more objective assessment method on teaching effectiveness that also reflects the impacts of context-based learning, we propose a theoretical approach from a unique perspective that recognizes teaching effectiveness as a result of the interplays between teacher, student, and context. The approach can be used to compute as well as to predict teaching effectiveness using machine and deep learning technologies, which brings strategical benefits to institutional management. In addition, we install into the approach a mechanism using tokens as incentives to assure the quality of subjective data input. The application framework for the approach is proposed leveraging blockchain. Each implementation of the framework by an establishment is a decentralized application that runs on its chosen blockchain. It is envisioned that the implementations together will form a collective ecology on context-based relative teaching effectiveness, which has the potential to fundamentally impact other academic practices besides teaching effectiveness measurement. The theoretical approach provides a common language to delineate teaching effectiveness from the context-based relative perspective and is customizable during implementation. The teaching effectiveness assessment using the approach downplays the roles played by bias (subjectivity) and hence is more objective than that by traditional student-oriented course evaluation.

KEYWORDS

Relative teaching effectiveness; Data science application; Blockchain; Context-based learning; Token economy

1 Introduction

One important mission at various teaching institutions is to assess teachers' abilities to teach courses. Results from the assessments are used to elicit useful data that indicates how teaching can be improved. The majority of teaching institutions worldwide currently employ

*Corresponding Author: heying@tjnu.edu.cn

course evaluation, a direct assessment by students, as the de-facto assessment/measurement of teaching effectiveness. It is useful and often can deliver insights on how teaching can be improved. However, this practice has been widely and continuously criticized, as many teachers and educational researchers believe that it is not as fair and accurate as it is supposed to be. Many factors that are not directly related to teaching effectiveness are more than often playing important roles in students' evaluations. In addition, recent educational research begins to explore the relationship between the composition of student body and teaching effectiveness, and it was shown that there exists a certain causal relationship between the two. Moreover, context-based teaching and learning also drew attention because of their positive impacts on student learning. Therefore, we try to approach the measurement of teaching effectiveness from a perspective that recognizes teaching effectiveness as a result of the interplays between teacher, student, and context. We develop a theoretical approach to model teaching effectiveness from that perspective and provide a measurement method for it. Leveraging machine and deep learning technologies, we can also build a software that simulates teaching by reusing historical records. An important feature of the software is that it can be used to predict teaching results (effectiveness), and hence it brings strategical benefits to institutional management. We augment the approach by installing into it a quality assurance mechanism based on the concept of token economy to address the issues related to subjectivity. The mechanism encourages conscientious subjective data inputs by using tokens as incentives. The application framework for the approach is also proposed based on blockchain and decentralized application technologies. In addition, we envision that the implementations of the framework by different establishments will help build a collective ecology on teaching effectiveness from our new perspective and it has the potential to fundamentally impact other academic practices.

We organize the rest of the paper as follows. Section 2 presents the theoretical background for this research. Section 3 presents our research motivations. Section 4 proposes our methodology for context-based relative teaching effectiveness. Section 5 proposes the application framework for it and envisions the collective ecology on context-based relative teaching effectiveness. Section 6 discusses our research and concludes this paper.

2 THEORETICAL BACKGROUND

In this section, we provide three literature reviews that qualify our research. Our research questions are given at the end of this section.

2.1 Token Economy and Crypto–Economy

A token economy is a management system using tokens to reinforce targets' positive behaviors. It originated from the fields of pedagogy and psychology (Kazdin, 1977; Phillips et al., 1971; Teodoro & Nathan, 1968). Token economy research prospered since the concept of blockchain was brought by Bitcoin (Satoshi, 2008). The research momentum on the relationship between tokens and digital financial systems continued when Ethereum was brought online in 2015 (Vitalik, 2014). Ethereum is a general-purpose decentralized application system running on its public blockchain, which is an essential feature that differentiates it from Bitcoin. Some recent research on token economy focused on promoting students' positive behaviors using tokens (Spiller et al., 2019). Most others focused on its business and financial implications (such as Teichmann & Falker, 2020; Santos et al., 2020; Lee, 2019).

Swiss Financial Market Supervisory Authority classifies tokens into three categories, whose counterparts in real life are currency, voucher for service, and stock share (FINMA, 2018). A token economy in which tokens are used as crypto-currency is defined as a crypto-economy (Cong et al., 2021; Kampakis, 2018; Kim & Chung, 2019; Lawrence et al., 2018). A crypto-economy has three major advantages over a traditional economy. First, in a traditional economy the created values by users are usually monopolized by a few giant establishments such as big companies and governments, and, in comparison, in a crypto-economy the created values by users are distributed more evenly to individual users. Consequently, crypto-economy is more attractive to individuals. This feature helps it enable decentralized and autonomous communities by creating a balance on growth. Second, in a crypto-economy, tokens transactions are almost costless, which enables individuals to gain more shares on the values created. Third, in a crypto-economy, token value will be augmented when its user base grows and when more desirable behaviors of its users are observed. Therefore, in this way, the tokens as incentives help a crypto-economy optimize its users' behaviors.

2.2 Educational and Teaching Assessment

Educational assessment has been defined as a systematic process of documenting and analyzing empirical data to improve student learning (Brookhart & Nitko, 2018; Mary, 2003; Nelson & Dawson, 2014). The assessment method can either be a direct one based on straight interpretations of data or an indirect one based on references on data. Educational assessment methods fall into three categories. (Suskie, 2018; Parker et al., 2013). First, a placement assessment is used to set up a baseline for a student to measure this/her growth. Second, with a formative assessment, an educator delivers feedback to students to help them learn better. Third, in a summative assessment, scores are placed on graded works. Teaching effectiveness assessment is a subtype of educational assessment, which is also known as teaching quality assessment, which belongs to the summative assessment category. It is also usually considered as a component of teacher quality assessment, which qualifies a teacher on a range of variables including "teaching effectiveness" that has been traditionally measured in terms of student achievement gains (Darling-Hammond & Youngs, 2002; Hanushek, 1971; Micheal, 2011).

Currently, course evaluation is the major teaching effective assessment practice employed worldwide. It is a direct assessment by students on their instructors' quality of teaching. The assessment criteria on a course evaluation usually include student engagement, clarity of instructions, fairness in grading, communication skills, presentation skills, teaching enthusiasm, etc. (Dunegan & Hrivnak, 2003; Kim et al., 2000; Rahman, 2006; Tang, 1997). Course evaluation is useful at yielding insights on how teaching can be improved from a student's perspective. However, it has been long and widely criticized as not being fair and accurate (Carrell & West, 2010; Dunegan & Hrivnak, 2003; Emery et al., 2003; Felton et al., 2003; Graham et al., 2020; Gray & Bergmann, 2003; Hamermesh & Parker, 2005; Isely & Singh, 2005; Krautmann & Sander, 1999; Langbein, 2008; Marks, 2012; McPherson et al., 2009; Merritt, 2007; Platt, 2010; Schaff & Siebring, 1974; Armstrong, 2011; Weinberg et al., 2009). In addition, Schaff has shown that the majority of teachers do not regard course evaluation as a true measurement of teaching effectiveness in a previous study (Schaff & Siebring, 1974). It has been typically argued that the factors not related to teaching effectiveness usually play important roles on a course evaluation such as easiness of grading, instructor's personality, gender, political views, looks, and ethnicity, etc. It has also been observed that in teaching schools in the US and other countries where student satisfaction plays an important role in determining continuous employment, to gain favorable ratings, teachers would teach contents in a way that is most caring to the slowest learners. Some other compromises in teaching to please students were also observed. In such situations, course evaluation dictates the quality of teaching. The contentions on course evaluation and other related issues ascended to the public domain (Anagnostopoulos et al., 2021).

Meanwhile, it has been shown that using a mathematical framework for assessing teaching effectiveness was useful, but it also has been proved to be ineffective at handling not-mathematics-specific concepts (Litke et al., 2021). A common language delineating teaching effectiveness to enable synergic collaboration on researching teaching quality was also searched for (Charalambous & Praetorius, 2020). Finally, reflections on teaching effectiveness stimulated researchers to rethink its definition. An empirical study demonstrated that teaching quality also depends on the composition of the student body (Fauth et al., 2021).

2.3 Context–Based Learning

Context-based learning is a relatively new field of study, and it refers to the practices of placing course content within a meaningful context to aid student learning. (Rose, 2012; Overton, 2016) There are two kinds of context, the social context of the learning environment and the real and concrete context of knowing. It has been shown by field studies that contexts play important roles in knowledge acquisition and processing, especially in the domains of science and technology education (Gramm et al., 2012; Kwon et al., 2015; Menthe & Parchmann, 2015; Taconis et al., 2016; Yamaguchi et al., 2020). Research on context-based learning focuses on constructing, for course content, a positive learning environment that is conducive to student learning the content (de Putter-Smits et al., 2013). By a semi-structured interview with computer science teachers, it was revealed that the majority of teachers had favorable views on context-based learning (Nijenhuis-Voogt et al., 2018). Currently, at many teaching institutions in the world, teachers are encouraged to adopt context-based teaching techniques such as the usage of multi-media, in-class experiments, and onsite study tours to help students learn better. It is becoming a global trend for teachers to use a variety of contexts to aid student learning.

3 MOTIVATIONS

First, we try to find a more objective perspective on teaching effectiveness assessment than student orientation and patch it with the concept of context-based learning since its practices are becoming more and more popular at teaching institutions. Consequently, we work on a unique perspective on teaching effectiveness that recognizes teaching results (effectiveness) as a consequence of the interplays between teacher, student, and context. From this perspective, we develop a theoretical approach to compute teaching effectiveness. This perspective is more objective than student orientation in that from this perspective teaching effectiveness is semi-formally calculated and the calculation downplays the roles played by biases (subjectivity).

Second, we try to develop a modeling approach as a common language to delineate teaching effectiveness from the new perspective (Charalambous & Praetorius, 2020) and at the same time avoid using a full mathematics-specific framework since it cannot handle educational concepts that are not mathematics-specific (Litke et al., 2021). Consequently, we employ a semi-formal modeling approach based on basic linear algebra and discrete mathematics concepts and operations. We avoid using fully formal mathematics-specific frameworks to delineate teaching effectiveness from a new perspective. The modeling approach is theoretical in nature, and it is intended to be a common foundation on which revisions and specializations can be made easily.

Third, by the semi-formal nature of our approach, we have recognized that it is impossible to rule out the issues related to subjectivity, as value assignments to some models used in our approach are still made subjectively. Therefore, we design and deploy a token economy (crypto-economy) based mechanism to assure the quality of such subjective data input referencing the relevant mechanisms by (Kim & Chung, 2019; Shibano et al., 2020).

Finally, we try to find the relevant data technologies to enhance our approach so as to make it feasible and efficient its application. As a result, we chose to use machine and deep learning technologies to enhance our approach application-wise. By doing so, we can extract data from historical records and apply the technologies to the data (for training) to facilitate the application of our approach. Without AI technologies, the historical records, the most important source of data for our approach, would not be reusable. In addition, considering the possible scenarios where our approach is implemented with regard to issues such as privacies, regulations, and competing interests, we envision that each implementation of our approach should be a blockchain application, because blockchain is flexible in addressing the issues and it is also a de facto foundation on which a crypto-economy is currently built.

4 METHODOLOGY: CONTEXT–BASED RELATIVE TEACHING EFFECTIVENESS

4.1 The Modeling Approach

First, we consider a course (C) as a collection of learning outcomes. In our approach, the concept of student (S) is modeled as a collection of properties, and so is the concept of context used by a course C (X_C).

$$C = (o^1, o^2, \dots, o^n), o^i \text{ is a learning objective of } C.$$

$$S = (p^1, p^2, \dots, p^l),\ p^i \text{ is a property of } S.$$

$$X_C = (q^1, q^2, \dots, q^m),\ q^i \text{ is a property of } X_C.$$

They can also be represented as vectors.

$$C^T = \begin{bmatrix} o^1 \\ o^2 \\ \cdot \\ \cdot \\ \cdot \\ o^n \end{bmatrix}, \qquad S^T = \begin{bmatrix} p^1 \\ p^2 \\ \cdot \\ \cdot \\ \cdot \\ p^l \end{bmatrix}, \qquad {X_C}^T = \begin{bmatrix} q^1 \\ q^2 \\ \cdot \\ \cdot \\ \cdot \\ q^m \end{bmatrix}$$

When a teacher is teaching a student, using a specific context, on a learning outcome of a course, he/she uses a specific subset of the student's and the context's properties.

$$S_O = (p^1, p^2, \dots, p^k),\ p^i \text{ is a property of } S \text{ and is relevant to the learning outcome } O \text{ of } C.$$

$$X_{C_O} = \left(p^1, p^2, \dots, p^j\right),\ p^i \text{ is a property of } X_C \text{ and is relevant to the learning outcome } O \text{ of } C.$$

We consider the teaching result for a learning outcome as an output produced by a triplet, which includes a teacher's teaching mode for the learning outcome (representing the teacher's ability to teach for the learning outcome), a collection of relative student-properties (representing a student), and a context (representing the context used by the teacher to facilitate teaching for the learning outcome).

$$R_O^{T\to S} = M_O^T(S_O, X_{C_O}),\ \ where \begin{cases} M_O^T: T's\ teaching\ mode\ for\ the\ learning\ outcome\ O \\ R_O^{T\to S}: the\ degree\ of\ completion, and\ R_O^{T\to S} \in [0,1] \end{cases}$$

Therefore, the collective teaching results, composed of the teaching results for all the learning outcomes, by a teacher T for a single student S on a course C is defined as follows.

$$SinRes_C^{T\to S} = \begin{bmatrix} R_{o^1}^{T\to S} \\ R_{o^2}^{T\to S} \\ . \\ . \\ . \\ R_{o^n}^{T\to S} \end{bmatrix}$$

Similarly, for a class of student S, the collective teaching results is defined as follows and it is in a matrix form.

$$ClassRes_C^{T\to \boldsymbol{S}} = (SinRes_C^{T\to S^1}, SinRes_C^{T\to S^2}, \ldots,\ SinRes_C^{T\to S^k}) = \begin{bmatrix} R_{o^1}^{T\to S^1} & R_{o^1}^{T\to S^2} & . & R_{o^1}^{T\to S^k} \\ R_{o^2}^{T\to S^1} & R_{o^2}^{T\to S^2} & . & R_{o^2}^{T\to S^k} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ R_{o^n}^{T\to S^1} & R_{o^n}^{T\to S^2} & . & R_{o^n}^{T\to S^k} \end{bmatrix},\ \ where\ k = |\boldsymbol{S}|$$

The effectiveness of teaching can be obtained based on the above definitions of teaching results. For example, suppose that the weights of a course's learning outcomes are prescribed for computation of teaching effectiveness, which is a usual practice at a teaching institution.

$$W_C{}^T = \begin{bmatrix} w^1 \\ w^2 \\ . \\ . \\ . \\ w^n \end{bmatrix},\ \ where \begin{cases} w^i: the\ weight\ of\ o^i\ of\ C \\ w^i\ \in [0,1]\ and\ \sum_i w^i = 1 \end{cases}$$

The corresponding teaching effectiveness for a single student can be defined as the sum of those weighted results.

$$SinEF_C^{T\to S} = (SinRes_C^{T\to S})^T \cdot W_C{}^T = (R_{o^1}^{T\to S} \cdot w^1\ +\ R_{o^2}^{T\to S} \cdot w^2, \ldots, R_{o^n}^{T\to S} \cdot w^n) = AbilityToTeach_C^T(S, X_C)$$

The collective teaching effectiveness for a class of students S can be defined as their average teaching effectiveness, which also signifies the teacher's ability to teach for the course.

$$ClassEF_C^{T\to \boldsymbol{S}} = \frac{\sum_S SinEF_C^{T\to S}}{|\boldsymbol{S}|} = AbilityToTeach_C^T(\boldsymbol{S}, X_C),\ \ \ where\ S\ \in \boldsymbol{S}$$

4.2 Comparison of Teachings Abilities

A teacher's ability to teach for a course is of our concern. It is not a single value that can be directly compared with one another. It is indeed a function. When comparing teachers' abilities to teach for the same course, effectiveness-wise, we need to apply their related teaching modes to the same class of students as well as the same context to obtain the results for comparison. In other words, in our approach comparing abilities to teach without target audience and context is meaningless.

$$CompareAbilitiesToTeach(T^1, T^2, \boldsymbol{S}, C, X_C) = Diff(ClassEF_C^{T^1\to \boldsymbol{S}}, ClassEF_C^{T^2\to \boldsymbol{S}})$$

Using the above definition, we argue that such context-based relative evaluation on teaching effectiveness is more objective than traditionally student-oriented one, as it is calculated using practical settings rather than is unilaterally and subjectively assigned by one party.

4.3 Training a Teaching Mode Simulation Engine

When relevant data on teaching effectiveness, student-properties, and context are available, a teacher's abilities to teach for a course can be obtained as a software, which simulates teaching in that mode, using machine-learning-related techniques on data training. (Ian et al., 2016; Murphy, 2012) Modern AI-based application frameworks, such as PyTorch and TensorFlow, can be used for this training purpose. (Aurélien, 2019; Ian, 2019)

A training data piece is of the following format.

$$TrainingDataPiece_C^T = (SinEF_C^{T\to S}, S, X_C)$$

Therefore, with a set of training data pieces, a teaching mode simulation software can be built.

$$Sim\_AbilityToTeach_C^T = Training(\underline{\boldsymbol{TrainingDataPiece_C^T}})$$

The software can be used to simulate a teacher's ability to teach a course (in the corresponding teaching mode). It outputs the predicted teaching effectiveness for an input in the format of either (S, X_C) or (S, X_C).

$$Sim\_SinEF_C^{T\to S} = Sim\_AbilityToTeach_C^T\left(\underline{S}, \underline{X_C}\right)$$

$$Sim\_ClassEF_C^{T\to \boldsymbol{S}} = Sim\_AbilityToTeach_C^T(\underline{\boldsymbol{S}}, \underline{X_C})$$

One thing to note is that when teaching a course, a teacher is supposed to use the same context for all of his/her students. Therefore, the values of X_C are supposed to be the same for all the students in a class..

When respective teaching simulation software can be obtained for two teachers on the same course. They can be employed to compare the two teachers' abilities to teach that course for a group of student using a same set of context.

$$CompareAbilitiesToTeach(T^1, T^2, \boldsymbol{S}, C, X_C) = Diff(Sim\_AbilityToTeach_C^{T^1}(\underline{\boldsymbol{S}}, \underline{X_C}), Sim\_AbilityToTeach_C^{T^2}(\underline{\boldsymbol{S}}, \underline{X_C}))$$

4.4 Benefits to Institutional Management

The teaching mode simulation software can be used to predict teaching effectiveness, which benefits institutional management in at least the following aspects.

4.4.1 Instructor Assignment

To maximize the teaching effectiveness for a class of student S on a course C, we can select the instructor who has the best predicted teaching effectiveness.

$$\hat{T} = \underset{T}{arg\,max}\ Sim\_ClassEF_C^{T\to \boldsymbol{S}}$$

To maximize the teaching effectiveness for a number of courses (for example, those offered in a single semester), we can select the instructor assignment permutation that generates the best cumulative effectiveness.

$$\hat{P} = \underset{P}{arg\,max} \sum_C \theta_C Sim\_ClassEF_C^{T(P)\to \boldsymbol{S}(C)},\ \ where \begin{cases} P \in \begin{Bmatrix} assignment \\ permutations \end{Bmatrix} \\ T(P): C's\ instroctor\ by\ P \\ \theta_C: C's\ weighting\ factor \\ \boldsymbol{S}(C): students\ in\ C \end{cases}$$

4.4.2 Context Resources Arrangement

A special case is that when there are conflicts in arrangements of contexts for courses (for example, some contexts are requested by the instructors for different courses but can only be offered to one of them), the previous selection shall be modified. In such case, an arrangement must specify an instance of instructor assignments as well as context assignments to courses.

$$\hat{P} = \underset{P}{arg\,max} \sum\nolimits_{C} \theta_C Sim\_AbilityToTeach_C^{T(P)}(\underline{\boldsymbol{S}(C)}, \underline{X_C(P)}), \ \ where \ \underline{X_C(P)} \ is \ C's \ context \ assigned \ by \ P$$

4.4.3 Course (Section) Registration

When a student S is deciding on which section of a course C to register, he/she can choose the section (and instructor equivalently) that generates the best predicted effectiveness for him/her.

$$\hat{T} = \underset{T}{arg\,max} \ Sim\_SinEF_C^{T \to S}$$

In addition, the student can also use the predicted teaching effectiveness to adjust his/her expectations on the course (or a section of it) and hence be mentally prepared.

4.5 Data Acquisition and Formatting

To build the teaching mode simulation software, we need to obtain historical records on teaching result, context, and student-properties. The data collected from the records will also need to be formatted, and the formats should be customized institution-wise. For example, to collect data on the teaching effectiveness on a course, we can map students' scores on graded items (obtained from historical records) into the teaching results for different learning outcomes for that course. Afterwards, using the course's weighting of its learning outcomes, we can calculate the final teaching effectiveness for that student on that course. Similarly, once the formats of context and student-properties (for different courses) are defined by an institution, relevant data can also be retrieved from historical records. The definitions on context and student-properties might be different among institutions, for they may have different understandings or viewpoints. Therefore, the definitions should be specified by the institution using it.

4.6 Quality Assurance for Subjective Data Inputs

During data collection for training the teaching simulation software, though the values for many data items can be obtained directly from records, there may exist some data items the values for which cannot be obtained objectively. For example, if "motivation" is defined as a property of a student taking a specific course. It is difficult to assign the value for it using direct data on the records. Such value may need to be assigned subjectively by some parties. For such difficult-to-evaluate data items, subjective inputs are unavoidable. Since we also aim to assure the qualities of such subjective data inputs, we propose to use a token-economy based mechanism (a peer review technology in the Library Science) to achieve this goal. Specifically, we apply a token economy for the evaluations of such data items. The economy encourages conscientious peer reviewers and punishes irresponsible peer reviewers by using tokens as incentives. The overview of such an economy is given in Figure 1.

The overall incentive mechanism using tokens is borrowed from STEEMIT (Kim & Chung, 2019). Each user in the economy can evaluate a data item when the evaluation is opened. Afterwards, the users can also peer-review each other's evaluation. Reward tokens are generated for the user of both an evaluation and a peer-review. For peer-review, a user can either down-vote or up-vote an evaluation. A user whose evaluation receives more up-votes than down-votes are rewarded with more tokens consequently. A user whose evaluation generally receives down-votes may lose tokens as a result. In addition, the users can also sell surplus tokens on a trading platform to make a financial gain.

Referring to the price stabilization mechanism in Shibano's work (Shibano et al., 2020), a policing unit is set up in the economy to control inflation by adjusting the rate of reward to-

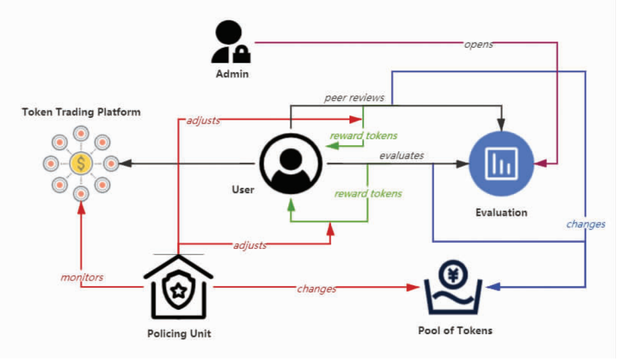

Figure 1 The overview of the token economy for quality assurance of subjective data inputs.

ken generation to guard against token-value manipulations as what can be done in the real-world financial markets. The other important objective of the policing unit is to make sure that the users "stay on the right track" by not placing financial gains as their top priority for participating in the economy. In addition, the policing unit will guard against the token value manipulations such as what can be done in the financial markets in reality.

The economy assures the quality of subjective data inputs by encouraging conscientious evaluation. For the economy to work in the long run, there shall also be a policy installed at the institutions that adopt the economy. For example, allocating a portion of the salaries or tuitions of the partakes (who are the economy users) to fill their accounts in the economy before the teaching evaluation period in each semester. After the evaluation is closed, the users can sell the tokens in their accounts to realize a financial gain or loss. Such policy would sustain value generation in the economy and is essential for its long-term continuous functioning.

The core design components of the framework are illustrated in Figure 2.

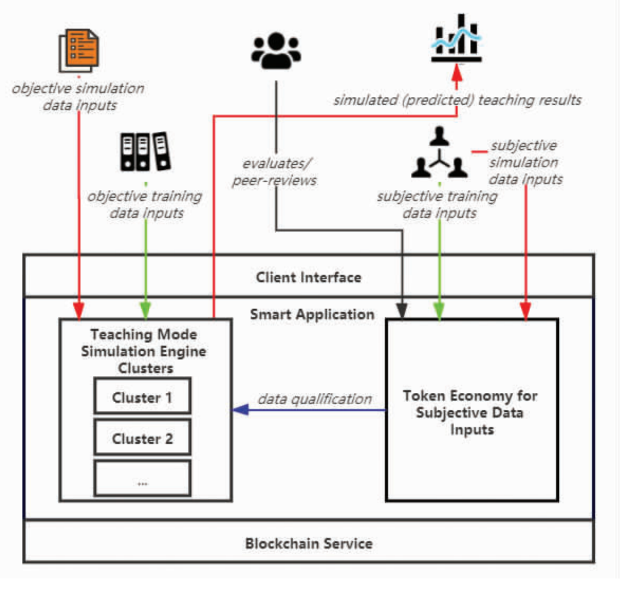

Figure 2 The overview of the application framework for context-based relative teaching effectiveness.

5 VALIDATION

In this section, we provide a case study for validation of our proposed method for computing and predicting context-based relative teaching effectiveness. Specifically, we will present a case study to compute a teaching simulation engine, use the engine to predict teaching effectiveness, and compare the predicted results with the real results.

We fetched the data of 12 Software Engineering course sections in 4 semesters from 2017 to 2021 and the academic records of 588 students in the 13 sections taught by the same instructor at Tianjin Normal University.

5.1 Gathering and Normalizing Data

To calculate the teaching simulation engine, we need to obtain historical records for teaching results $\underline{R}^{S}_{TC}$ (for $ClassEF_C^{T\rightarrow S}$) and those for student properties $\underline{S}$.

5.1.1 Gathering and Normalizing Teaching Results

$\underline{R}^{S}_{TC}$ is a collection of historical teaching results data for a set of students $\underline{S}$ for the course C. Each piece of $\underline{R}^{S}_{TC}$ is $\underline{R}^{S}_{TC}$, which are the observed teaching results of a student $\underline{S}$ for the course C in the past, where $\underline{R}^{S}_{TC} = (\underline{q}^1, \underline{q}^2, \ldots, q^n)$, as shown and calculated in below.

$$\underline{R}^{S}_{TC} = \begin{bmatrix} \underline{q}^1 \\ \underline{q}^2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{q}^n \end{bmatrix}^T \qquad \underline{q}^i = \sum_k \left(a_{i_k} \cdot p_{i_k}\right) \qquad \underline{q}^i \in [0,1]$$

p_{i_k}: k^{th} *component of S's performace score on* o^i *of C*

a_{i_k}: *coefficient of* p_{i_k}

A component of $\underline{R}^{S}_{TC}$ is q^i, a student's observed performance on o^i. It is calculated based on the student's observed performances on related deliverables.

For example, assume that there are two sub-components for a learning objective o^i. The first one is the performance on a homework problem with a full score of 5 and a coefficient of 0.4. The second one is the performance on an exam question with a full score of 12 and a coefficient of 0.6. One student's actual performances on these two sub-components are 3 and 10. Therefore, his overall performance on this learning objective is calculated as follows.

$$0.4 \cdot (3/5) + 0.6 \cdot (10/12) = 79\%$$

5.1.2 Gathering and Normalizing Student Properties Data

Each piece of $\underline{S}$ is $\underline{S}$, where S = $(p^1, p^2, \ldots, p^m)$. It is an instantiated student model whose properties' values are calculated based on the (observed) data collected from this student's record before his/her taking the course. It is shown as follows.

$$\underline{S}^T = \begin{bmatrix} \underline{p}^1 \\ \underline{p}^2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{p}^m \end{bmatrix} \qquad \underline{p}^i = \sum_k \left(b_{i_k} \cdot h_{i_k}\right) \qquad \underline{p}^i \in [0,1]$$

h_{i_k}: k^{th} *component of S's evalution score on the* i^{th} *property*

b_{i_k}: *coefficient of* h_{i_k}

5.2 Model Configurations

We provided the following configuration for a Software Engineering course, which encompasses 16 teaching objectives, consulting the Software Engineering Body of Knowledge (SWEBOK), as shown in Table 1.

TABLE I CONFIGURATION OF A SOFTWARE ENGINEERING COURSE

Software Engineering Course Objectives	
ID	Objective Category
o^1	Software Crisis
o^2	Software Definition
o^3	Software Life Cycle Models
o^4	Requirements Analysis
o^5	Object–Oriented Programming
o^6	Software Architecture Design
o^7	Software Detailed Design
o^8	UML
o^9	Software Construction
o^{10}	Software Testing
o^{11}	Software Maintenance
o^{12}	Software Project Management
o^{13}	Software Project Documentation
o^{14}	Software Configuration Management
o^{15}	Software Engineering Code Of Ethics
o^{16}	Project–Based Software Development Experiences

Additionally, based on our experiences, we provided the following configuration for the student properties required by the Software Engineering course, as shown in Table 2.

TABLE II CONFIGURATION OF A SOFTWARE ENGINEERING STUDENT

Software Engineering Student Properties		
ID	Explanation	Property Category (Source Course)
p^1	Mathematical and Analytical Capability	College Algebra I
p^2		College Algebra II
p^3		Discrete Mathematics
p^4	Computer Fundamentals	Operating System
p^5		Computer Organization and Architecture
p^6	Programming Experiences	C++ Programing Or Java Programming
p^7	Technical Writing and the Abilities to Express and Understand	National College Entrance Examination on Chinese Literature and Language
p^8	Capability of Teamwork	Class Monitor´s Evaluations in their Profile Archives

When calculating the final course grade using performance scores on each of the course objectives, we need to take into consideration the weights of the objectives, which, according to the tradition of teaching the course, is given in Table 3.

TABLE III WEIGHTS DISTRIBUTION OF THE SOFTWARE ENGINEERING OBJECTIVES

Weights Distribution of the Course Objectives		
ID	Objective Category	Percentage
o^1	Software Crisis	2
o^2	Software Definition	2
o^3	Software Life Cycle Models	10
o^4	Requirements Analysis	10
o^5	Object–Oriented Programming	15
o^6	Software Architecture Design	2
o^7	Software Detailed Design	2
o^8	UML	15
o^9	Software Construction	10
o^{10}	Software Testing	2
o^{11}	Software Maintenance	2
o^{12}	Software Project Management	2
o^{13}	Software Project Documentation	2
o^{14}	Software Configuration Management	2
o^{15}	Software Engineering Code Of Ethics	2
o^{16}	Project–Based Software Development Experiences	20

In addition, at TJNU, course final grade is given on a 100-points scale. However, we argue that argue qualitative grades in letters are more comparable. As a result, we adopted the letter-based grade system using a conversion shown in Table 4.

TABLE IV PERCENTILE TO LETTER GRADE CONVERSIONS

Percentile to Letter Grade Conversions										
Prct.	100–93	92–90	89–87	86–84	83–80	79–75	74–70	69–66	65–60	59–0
Letr.	A	A–	B+	B	B–	C+	C	C–	D	F

A problem occurred when we were collecting the data for p7 in Table 2, as the national college entrance exams for it use different digital grade systems (either 150 nation-wide or 120 based for Jiangsu Providence). Therefore, we provided a data normalization as given in Table 5.

TABLE V PERCENTILE TO LETTER GRADE CONVERSIONS FOR P^7

Percentile to Letter Grade Conversions for P^7									
Students not from Jiangsu province									
150–140	139–135	134–131	130–126	125–120	119–113	112–106	105–99	98–90	89–0
A	A–	B+	B	B–	C+	C	C–	D	F
Students from Jiangsu province									
120–112	111–108	107–105	104–101	100–96	95–90	89–84	83–80	79–72	71–0
A	A–	B+	B	B–	C+	C	C–	D	F

For demonstration's purpose, we only chose the simple context information for the course, online or in-class. The course has historically been taught in-class. However, during the Covid-19 period, the course was taught online for the last two consecutive semesters.

TABLE VI SOFTWARE ENGINEERING COURSE CONTEXT

Context	In–class
	Online

5.3 Strategy

We first computed the teaching simulation engine for the instructor using the In-class context for the 7 sections of students (317 in total) between 2017 and 2019, of which 5 sections were used for training. The last two sections were used for testing the engine. Afterwards, we computed the teaching simulation engine for the instructor using the Online context for the 6 sections of students (271 in total) between 2020 and 2021, of which 4 sections were used for training. (Due to the impacts of COVID 19, most of the classes from 2020 till the Spring of 2021 were offered online. However, the final exams were conducted in-class.) The last two sections were used for testing the engine. Then we used the two different engines to predict teaching results for the other context to see whether context-based learning played a difference in prediction of teaching effectiveness.

5.4 Results

5.4.1 Results for In–class Context Mode

We used two-layer feed-forward neural networks and the classical BP algorithm (with learning rate η = 0.2) for training the engine. The training data set is composed of 230 data pieces (5 sections between 2017 to 2019), which were collected from 5 sections of the course. A training data piece is in the format of ($\underline{S}$, $\underline{G}$), and an example of it is shown below.

TABLE VII A TRAINING DATA PIECE EXAMPLE

A Training Data Piece Example									
$\underline{S}$									$\underline{G}$
ID	p1	p2	p3	p4	p5	p6	p7	p8	
Val.	B+	A–	B+	C+	A	B–	B	B–	B

Once we had the engine trained, we used two testing data sets, which are composed of totally 87 data pieces that were collected from 2 sections in 2019, to test its predication accuracy. Specifically speaking, for any testing data piece ($\underline{S}$, $\underline{G}$), we input ($\underline{S}$) into the engine, which would produce a prediction result ($\underline{G}$'), which will be compared to $\underline{G}$ in the piece for accuracy. The overall predication results are shown as follows.

TABLE VIII PREDICTION RESULTS BY THE SIMULATING ENGINE FOR IN-CLASS CONTEXT MODE

Prediction Results by the Teaching Simulating Engine					
Differences	P_{Crct}	P_{Half}	P_{One}	P_{Err}	Data Pieces Num.
	None	Half Letter	One Letter	More than One Letter	
Set 1	32	5	1	3	41
Set 2	34	6	3	3	46
Cumulative	66	11	4	6	87
Percentage	75.86	12.64	5.75	6.90	100

The engine's predication accuracy is 75.86%. We argue that a certain amount of error shall be allowed when calculating its credibility (effectiveness), as the predication of grade range is more meaningful and useful than the predation of exact grade to stakeholders. Therefore, we count the half-letter difference percentage rate into the engine's credibility, as given in below.

$$Credibility\ of\ the\ Teaching\ Simulating\ Engine\ (In-Class) = P_{Crct} + P_{Half} = 88.5\%$$

In this case, the engine's credibility is 88.5%. In other words, there is 88.5% chance that a student's grade will be in half of a letter grade distance from the engine's predicted result.

5.4.1 Results for Online Context Mode

We used the training data set that is composed of 181 data pieces (4 sections in 2020), which were collected from 4 sections of the course in 2020. The testing data set is composed of 90 data pieces (2 sections in 2021).

TABLE IX PREDICTION RESULTS BY THE SIMULATING ENGINE FOR ONLINE CONTEXT MODE

Prediction Results by the Teaching Simulating Engine					
Differences	P_{Crct}	P_{Half}	P_{One}	P_{Err}	Data Pieces Num.
	None	Half Letter	One Letter	More than One Letter	
Set 1	30	6	2	5	43
Set 2	37	4	4	2	47
Cumulative	67	10	6	7	90
Percentage	74.44	11.11	6.67	7.78	100

$$Credibility\ of\ the\ Teaching\ Simulating\ Engine\ (Online) = 85.55\%$$

5.4.2 Results for " Cross" Verification

We also applied the two teaching effectiveness simulation engines to the testing data sets of the mode. The prediction results were shown in the following table.

TABLE X PREDICTION RESU LTS BY THE SIMULATING ENGINE FOR IN-CLASS CONTEXT MODE ON ONLINE MODE TESTING DATA SETS

Prediction Results by the Teaching Simulating Engine					
Differences	P_{Crct}	P_{Half}	P_{One}	P_{Err}	Data Pieces Num.
	None	Half Letter	One Letter	More than One Letter	
Set 1	15	18	2	8	43
Set 2	19	10	8	10	47
Cumulative	34	28	10	18	90
Percentage	37.78	31.11	11.11	20	100

The credibility of the corresponding prediction is 68.89%.

TABLE XI PREDICTION RESULT S BY THE SIMULATING ENGINE FOR ONLINE CONTEXT MODE ON IN-CLASS MODE TESTING DATA SETS

Prediction Results by the Teaching Simulating Engine					
Differences	P_{Crct}	P_{Half}	P_{One}	P_{Err}	Data Pieces Num.
	None	Half Letter	One Letter	More than One Letter	
Set 1	17	10	5	9	41
Set 2	16	9	11	10	46
Cumulative	33	19	16	19	87
Percentage	37.93	21.84	18.39	21.84	100

The credibility of the corresponding prediction is 69.77%.

5.4.3 Discussions

As it was shown in Table 8 and 9, we argue that the prediction results were successful as each of their credibility is more than 85%. In other words, there is more than 85% chance that a student's grade will be in half of a letter grade distance from each of the engines' predicted results. Therefore, we argue that, with cautious optimism, our modeling and prediction approach for teaching effectiveness is useful.

Table 10 and 11 also show that the two contexts made a difference in the predictions' accuracies (credibilities), as predications using different contexts than the actual one yielded lower credibilities (both around 69%). Therefore, we argue that, even for the same instructor, context impacts a course delivery and such impacts are reflected in the corresponding teaching simulation engine built for the context. It is necessary to include context when modeling and predicting teaching effectiveness. Nevertheless, we can also perceive some degree of course delivery consistency in the corresponding prediction results, which shows that the impacts on prediction credibilities by the contexts were around -15% to -18%. In all, we argue that since such impacts were significant enough, it justifies our notion that contexts should be included in our teaching effectiveness modeling and prediction approach.

One thing to note is that when we were looking into the actual results (performance scores) on the tested items of the online sections, we found that they were comparatively and noticeably lower than those of the in-class sections. Our discussions gave many possible reasons such as the lack of concentration, logistics problems, and the inability to enforce self-disciplines, etc. Reasons or facts like these can be further studied and factored into our model in the future.

6 APPLICATION FRAMEWORK AND ECOLOGY FOR CONTEXT–BASED RELATIVE TEACHING EFFECTIVENESS

We also envision an application framework for context-based relative teaching effectiveness as a smart-contract-like application running on a public blockchain. An implemented framework together with its users can be considered as a sub-ecology for context-based relative teaching effectiveness.

6.1 Blockchain Service

A blockchain service runs at the bottom level of the framework. It provides a decentralized service infrastructure for the upper components. First, it provides distributed data storage for the whole framework. Second, it enables the core business logic of the framework to run as a decentralized smart application. The blockchain can be either a permissioned or permission-less one. An institution or a group of institutions that implement(s) the framework can choose an appropriate type of blockchain (private, consortium, or public) based on its/their preference(s).

6.2 Smart Application

A smart application is installed on the chain. It is a decentralized software application running on all nodes of the chain. It can either be like an Ethereum smart application (Antonopoulos & Wood, 2018) on a public blockchain or a restrictive application running on a consortium or private chain. It can also be built as the sole application running on a dedicated chain such as Bitcoin.

The application is composed of two parts. The first part is responsible for training and storing teaching mode simulation software components. The second part instantiates the token economy for subjective data inputs and provides a data qualification service (on subjective data inputs obtained from upper-level components) to the first part by using the economy.

6.3 Client Interface

A client interface is a front-end software installed at various end nodes of the chain. It is used to access the smart application. It accepts inputs from end-users and relays the inputs to the application. It also relays the application outputs to the users.

6.4 Training Teaching Mode Simulation Software Components

For training the simulation software components, user inputs training data through the client interface. The objective parts of the data are directly fed to the corresponding component in the smart application. The subjective parts, normally with values to be determined, are sent to the token economy part of the application. The values will be determined and qualified by the economy. Afterwards, the subjective data with qualified values are sent to the simulation software training part of the application, where both the objective and subjective data values are employed for training the relevant teaching mode simulation software component.

6.5 Obtaining Simulation Results

Once a teaching mode simulation software component is trained, it can be used to produce simulated teaching results (and effectiveness as well). A user sends the data for simulation inputs through the client interface to the smart application. The objective parts of the data are sent to the component directly. The values of the subjective parts of the data are determined and qualified by the token economy and then sent to the component, which outputs simulated teaching results (effectiveness) to the user through the client interface.

6.6 Toward an Ecology of Context–Based Relative Teaching Effectiveness

Various interest groups can have different implementations of the framework based on their diverse understandings of student-properties, contexts, and/or priorities of the learning outcomes of courses, which may result in variances in the results of the teaching effectiveness computation. To address this heterogeneity issue, we envision that each interest group sets up their own application of the framework and establishes necessary data exchange protocols if they want to exchange data with one other, as illustrated in Figure 3.

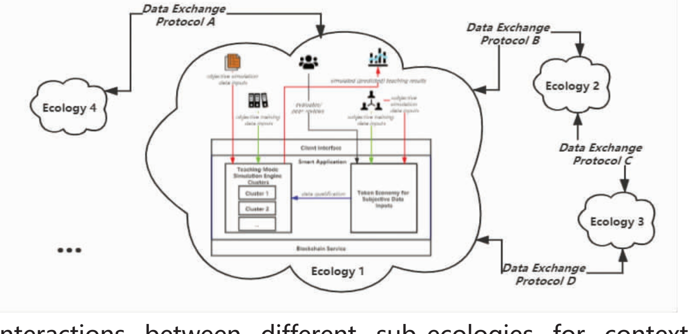

Figure 3 The interactions between different sub-ecologies for context-based relative teaching effectiveness.

An implementation of the framework together with the interest (user) group form a specific sub-ecology of context-based relative teaching effectiveness. Two such sub-ecologies can interact with each other by using a data exchange protocol that translates incoming data using their own interpretations. The sub-ecologies together form an overall ecology on context-based relative teaching effectiveness.

7 DISCUSSIONS AND CONCLUSIONS

In this paper, we proposed a theoretical modeling approach from a context-based relative perspective on teaching effectiveness. The perspective recognizes teaching results (effectiveness) as a consequence of the interplays between teacher, student, and context. We primarily aimed to provide a modeling approach to delineate teaching effectiveness from that perspective. The approach is theoretical and abstract in nature and hence is open to adaption.

To facilitate the application of our approach, we employed machine and deep learning technologies in order to extract data from historical records, which compose the major source of data for our approach.

We demonstrated the application of our approach using a case study that involved one course, one instructor, and 588 students in 13 sections from 2017 to 2021. Around half of the sections received the course in-class, and the other half received the course online due to the impacts of COVID 19. The demonstration yielded optimistic results, which shows that our approach is useful.

The semi-formal nature of our approach cannot rule out the issues related to subjectivity, for which we proposed to employ a token economy (crypto-economy) based mechanism for quality assurance on subjective data inputs.

Considering the possible scenarios where our approach can be adopted and implemented, we propose an application framework for it leveraging the blockchain technology, because blockchain is flexible in coping with the heterogeneous reality, in which issues such as privacy, regulation, and competing interests complicate the implementation of our framework. We envision that each interest group builds its own implementation of our framework based on its understandings and interpretations of the model elements used in our approach, and the implementation together with the interest group (users) form a sub-ecology on context-based relative teaching effectiveness. The sub-ecologies together form an overall ecology on context-based relative teaching effectiveness, in which a data exchange protocol shall be made between two communicating sub-ecologies.

Once the sub-ecologies are established, they potentially can have impacts on other academic routines. For example, when admitting a student to a course or a program, his/her relevant properties can be retrieved from the corresponding sub-ecology and are used to predict his/her potential of success in the course or the program. It helps institutional management in screening admission applicants. For another example, when a teacher is transferring to another institution, to evaluate his/her qualification and potential in teaching success in the new institution, his/her teaching mode simulation software component can be retrieved from his/her previous sub-ecology and are applied to the students in the target sub-ecology that the institution resides in, and the simulated teaching results can be used as references.

A contribution by our research is that it provides a more objective method than course evaluation for teaching effectiveness assessment by downplaying the roles played by bias (subjectivity) during the assessment, though due to the semi-formal nature of our approach, some subjective data inputs can still be present when using our approach. A token-economy-based mechanism is employed to assure the qualities of the subjective data inputs.

A side benefit of our approach is that, by using the simulated teaching results (effectiveness), institution management can conduct some academic arrangements more effectively, e.g. course scheduling, instructor assignments, and course (section) registration.

Since the major aim of our approach is to provide an overall abstract approach on context-based relative teaching effectiveness, it leaves the implementation details to the interested users or user group. For example, the definition on a context used for a course and the definition on the student properties referenced by a course. Different users can make their own definitions based on their understandings and the characteristics of their specific educational domains.

In addition, determining student-properties and the correspondences between them and the relevant courses is a challenge. This challenge illustrates the "experimental" aspect of this research. A forum should be established for field practitioners, academic researchers, and related education professionals to collaborate on doing the experiments synergically. Strategically speaking, we suggest using a bottom-up approach, in which the determinations concerning individual courses take precedence. As the knowledge deepens, the results are synthesized and refined to make the determinations on a higher level, e.g., track level, program level, school level, university level, etc. In the future, we would explore this direction.

Moreover, in the near future, we envision that storing student properties data on a blockchain may be prohibited by some regulations or laws for privacy concerns. Therefore, strategies and special regulations on such data usage shall be developed. The current encryption and authentication technologies should be more than enough to support them, e.g., encryption of student data by private and public keys set and data access by authorization technologies.

Finally, in our approach, teaching quality is measured in terms of "the completeness of learning outcomes". It can be extended to take more viewpoints into account. For example, we can extend the measurement to cover some other criteria such as "student satisfaction" and "engagement", which can be added directly to the course's learning outcomes, and teaching result (effectiveness) is updated accordingly by the computation.

We will work on addressing the above limitations for future work.

ACKNOWLEDGMENT

The work in this paper is jointly funded by Tianjin Municipality Natural Science Foundation (18JCYBJC44500) and the National Social Science Foundation of China (No. 20BTQ084).

AUTHOR CONTRIBUTIONS

Kun Tian: Conceptualization, Methodology, Writing-Original Draft, Funding acquisition, Supervision; Ying He: Resources, Writing - Review & Editing, Project administration; Xiaoran Xu: Data Curation, Investigation, Validation; Jiangyang Fu: Data Curation, Investigation, Validation.

DECLARATIONS OF INTEREST

None.

Reference

Anagnostopoulos, D., Wilson, S., & Charles–Harris, S. (2021). Contesting quality teaching: Teachers´ pragmatic agency and the debate about teacher evaluation. *Teaching and Teacher Education, 98*, 103246.

Antonopoulos, A. M., & Wood, G.(2018). *Mastering Ethereum: Building Smart Contracts and DApps* (1st ed). O´Reilly Media.

Armstrong, J. S. (2011). *Natural learning in higher education*. Retrieved Feb 5, 2021, from https://faculty.wharton.upenn.edu/wp–content/uploads/2014/09/Natural–Learning–in–Higher–Education_2.pdf.

Aurélien, G. (2019). *Hands–on machine learning with scikit–learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems* (2nd ed). O´Reilly Media.

Boudett,K. P., City, E. A., & Murnane, R. J. (eds). (2013). *Data wise, revised and expanded edition: A step–by–step guide to using assessment results to improve teaching and learning revised and expanded edition*. Harvard Education Press.

Brookhart, S. M., & Nitko, A. J. (2018). *Educational assessment of students* (8th ed). Pearson.

Carrell, S., & West, J.(2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy, 118*, 409–432.

Charalambous, C. Y., & Praetorius, A.–K. (2020). Creating a forum for researching teaching and its quality more synergistically. *Studies in Educational Evaluation, 67*, 100894.

Cong, L. W., Li, Y., & Wang, N.(2021). Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies, 34* (3), 1105–1155.

Darling–Hammond, L., & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically–based research" actually tell us?. *Educational Researcher* , *31* (9), 13–25.

de Putter–Smits, L. G. A., Taconis, R., & Jochems, W. M. G. (2013). Mapping context–based learning environments: The construction of an instrument. *Learning Environments Research, 16* (3), 437–462.

Dunegan, K. J., & Hrivnak, M. W. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education, 27* (3), 280–303.

Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: A critique of student evaluations of teaching effectiveness. *Quality Assurance in Education, 11* (1), 37–46.

Fauth, B., Atlay, C., Dumont, H., & Decristan, J. (2021). Does what you get depend on who you are with? Effects of student composition on teaching quality. *Learning and Instruction, 71*, 101355.

Felton, J., Mitchell, J. B., & Stinson, M. (2003). Web–based student evaluations of professors: The relations between perceived quality, easiness, and sexiness. *Assessment & Evaluation in Higher Education, 29* (1), 91–108.

FINMA.(2018). *Guidelines for enquiries regarding the regulatory framework for initial coin offerings(icos)*. Bern, Switzerland: FINMA.

Graham, L. J., White, S. L. J., Cologon, K., & Pianta, R. C. (2020). Do teachers´ years of experience make a difference in the quality of teaching? *Teaching and Teacher Education, 96*, 103190.

Gramm, A., Hornung, M., & Witten, H. (2012). Email for you(only?): Design and implementation of a context–based learning process on internetworking and cryptography. *Proceedings of the 7th Workshop in Primary and Secondary Computing Education*, Hamburg, Germany.

Gray, M., & Bergmann, B. R.(2003). Student teaching evaluations: Inaccurate, demeaning, misused. *Academe, 89* (5), 44–46.

Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors´ pulchritude and putative pedagogical productivity. *Economics of Education Review, 24* (4), 369–376.

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review, 61* (2), 280–288.

Ian, G., Yoshua, B., & Courville, A. (2016). *Deep learning*. The MIT Press.

Ian, P.(2019). *Programming pytorch for deep learning: Creating and deploying deep learning applications* (1st ed). O´Reilly Media.

Isely, P., & Singh, H.(2005). Do higher grades lead to favorable student evaluations? *The Journal of Economic Education, 36* (1), 29–42.

Kampakis, S.(2018). Why do we need tokenomics? *The Journal of the British Blockchain Association, 1*, 1–4.

Kazdin, A. E.(1977). *The token economy*. Springer.

Kim, C., Damewood, E., & Hodge, N. (2000). Professor attitude: Its effect on teaching evaluations. *Journal of Management Education, 24* (4), 458–473.

Kim, M. S., & Chung, J. Y. (2019). Sustainable growth and token economy design: The case of steemit. *Sustainability, 11* (1), 167.

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review, 18* (1), 59–63.

Kwon, C., Kim, Y., & Woo, T.(2015). Digital–physical reality game: Mapping of physical space with fantasy in context–based learning games. *Games and Culture, 11* (4), 390–421.

Langbein, L.(2008). Management by results: Student evaluation of faculty teaching and the mis–measurement of performance. Economics of Education Review, 27 (4), 417–428.

Lawrence, L., John, J., Harry, M., & Shaquile, N. (2018). The convergence ecosystem. Retrieved Feb 14, 2021., from https: // outlierventures. io/wp–content/uploads/2018/03/The_Convergence_Ecosystem_Report_Outlier_Ventures_2018.pdf

Lee, J.Y. (2019). A decentralized token economy: How blockchain and cryptocurrency can revolutionize business. *Business Horizons, 62* (6), 773–784.

Litke, E., Boston, M., & Walkowiak, T. A. (2021). Affordances and constraints of mathematics –specific observation frameworks and general elements of teaching quality. *Studies in Educational Evaluation, 68,* 100956.

Marks, P. (2012). Silent partners: Student course evaluations and the construction of pedagogical worlds. *Canadian Journal for Studies in Discourse and Writing/Rédactologie, 24*.

Mary, A. J. (2003). *Assessing academic programs in higher education*. Josei–Bass.

McPherson, M., Jewell, T., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal, 35*, 37–51.

Menthe, J., & Parchmann, I.(2015). Getting involved: Context–based learning in chemistry education. In M. Kahveci & M. Orgill (Eds.), *Affective dimensions in chemistry education* (pp. 51 –67). Springer Berlin Heidelberg.

Merritt, D. (2007). Bias, the brain, and student evaluations of teaching. *St. John´s Law Review, 82.*

Michael, S. (2011). *The highly qualified teacher*. Teachers College Press.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. The MIT Press.

Nelson, R., & Dawson, P. (2014). A contribution to the history of assessment: How a conversation simulator redeems socratic method. *Assessment & Evaluation in Higher Education, 39* (2), 195–204.

Nijenhuis –Voogt, J., Meijer, P. C., & Barendsen, E. (2018). Context –based teaching and learning of fundamental computer science concepts: Exploring teachers´ ideas. *Proceedings of the 13th Workshop in Primary and Secondary Computing Education,* Potsdam, Germany.

Overton, T. (2016). Context and problem–based learning. *New Directions in the Teaching of Physical Sciences*. 2007 (3),7–12.

Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1971). Achievement place: Modification of the behaviors of pre–delinquent boys within a token economy1,2. *Journal of Applied Behavior Analysis, 4* (1), 45–59.

Platt, M. (2010). What student evaluations teach. *Perspectives on Political Science, 22*, 29–40.

Rahman, K.(2006). Learning from your business lectures: Using stepwise regression to understand course evaluation data. *Journal of American Academy of Business, Cambridge, 9,* 272–279.

Rose, D. E. (2012). Context–based learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 799–802). Springer US.

Santos, D., Sanfins, M. A., Nacif, P. L., & Nascimento, A. (2020). The token economy assets for the blockchain era global 2020. *Global Media Journal, 19*, 1.

Satoshi, N. (2008). *Bitcoin whitepaper*. Retrieved Feb 14 from http://satoshinakamoto.me/whitepaper/

Schaff, M. E., & Siebring, B. R. (1974). A survey of the literature concerning student evaluation of teaching. *Journal of Chemical Education, 51* (3), 150.

Shibano, K., Lin, R., & Mogi, G. (2020). Volatility reducing effect by introducing a price stabilization agent on cryptocurrencies trading. *Proceedings of the 2020 The 2nd International Conference on Blockchain Technology*, Hilo, HI, USA.

Spiller, A. N., Caro, K., Garay, J. A., & Marcu, G. (2019). Supporting behavior management with a classroom display providing immediate feedback to students. *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare,* Trento, Italy.

Suskie, L. (2018). Assessing student learning: *A common sense guide* (3rd ed). Anker.

Taconis, R., den Brok, P., & Pilot, A. (2016). *Teachers creating context–based learning environments in science*. SensePublishers, Rotterdam.

Tang, T. L.–P. (1997). Teaching evaluation at a public institution of higher education: Factors related to the overall teaching effectiveness. *Public Personnel Management, 26* (3), 379–389.

Teichmann, F., & Falker, M.–C. (2020). Blockchain: Implications of the impending token economy. Scientific and Technical Revolution: Yesterday, Today and Tomorrow, Cham.

Teodoro, A., & Nathan, A. (1968). *The token economy: A motivational system for therapy and rehabilitation.* AppletonCentury Crofts.

Vitalik, B. (2014). *Ethereum whitepaper*. Retrieved Feb 9 from http://ethereum.org/en/whitepaper/

Weinberg, B. A., Hashimoto, M., & Fleisher, B. M. (2009). Evaluating teaching in higher education. *The Journal of Economic Education, 40* (3), 227–261.

Yamaguchi, K., Iwata, M., Vargo, A., & Kise, K. (2020). Mobile vocabulometer: A context–based learning mobile application to enhance english vocabulary acquisition. *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers,* Virtual Event, Mexico.