LearnEval Peer Assessment Platform: Iterative Development Process and Evaluation

Gabriel Badea and Elvira Popescu, *Senior Member, IEEE*

***Abstract*—Peer assessment is a valuable educational activity, which promotes learner engagement and critical thinking, while also decreasing the time spent by the teachers on evaluating students' work. Over the past decades, several platforms have been proposed to support the peer assessment process. The current landscape is dominated by special-purpose, context-specific peer evaluation systems, which were designed to support a particular type of learning activity, artifact, or discipline. In this article, we propose an alternative, a generic and comprehensive peer assessment platform, called LearnEval, which is very flexible and able to accommodate a variety of pedagogical scenarios. Distinctive functionalities include: highly configurable peer assessment workflow, several automatic and manual review allocation mechanisms, initial calibration/training options for the students, complex dashboard with statistics, scores and graphical visualizations, as well as an extensive open learner model. The platform was developed in an iterative manner, with several extensions and improvements informed by the findings of the initial pilot studies. The article also reports on a practical application of LearnEval in a project-based learning scenario, with very promising results; multiple perspectives of the peer assessment process are analyzed: student behavior, peer grades validity, and perceived learning experience. Comparative results with an earlier study show increased levels of system usability, grading validity, and student satisfaction.**

***Index Terms*—Peer assessment system, peer grading validity, peer reviewing, project-based learning (PBL), student satisfaction, system usability.**

I. Introduction

PEER assessment is an effective educational activity that has started to be adopted in many learning contexts, especially in the past decade. The literature provides plentiful evidence that peer evaluation offers many advantages for the main actors involved in the educational process [37]. On one hand, the approach benefits the student by means of encouraging critical thinking and more time spent on the task, providing access to more feedback (not only the one from the teacher) as well as access to different solutions for a given task, offering new perspectives [26], [36]. On the other hand, the approach benefits also the instructor, by means of scaling the grading task in large classes or massive open online courses (MOOCs) [30], [34], decreasing the time allocated for assessing solutions, and reducing the problem of "free-rider" students; this "free-rider" challenge exists in group tasks where some team members perform better than others and the instructor has difficulties in assigning distinct grades according to the particular contribution of each student [15].

Peer assessment has been employed in a wide range of topics and scenarios, but a particularly successful usage has been in classes where the output is creative and the evaluation can be highly subjective [39]. Despite its many advantages, the peer assessment process is not free of impediments. Some of the most important challenges include: students' reluctance to evaluate their peers, the friendship relationships between learners which could lead to subjective assessments, and "tit-for-tat" scoring (i.e., students' worries that assigning low grades might influence their peers to also assign low grades as a revenge) [13], [37].

The technological advancements and the need for a solution to the time-consuming, low scaling task of manually evaluating student work have led to the emergence of computer-supported peer assessment systems. The first such platforms were developed over two decades ago [15]. Over time, the affordances and complexity of the systems grew steadily to accommodate the demands and requirements of both instructors and learners. During the last two decades, dozens of peer assessment systems, each with its unique features, have emerged [18], and the current landscape is in a continuous growth. Whereas some systems focus on the assessment workflow and the steps involved, enhancing it in various ways [24], others focus on the affordances they offer to the instructors and learners [40].

In this context, it is important to provide a peer assessment platform that unifies and improves existing features from the current solutions to better fulfill educational stakeholders' needs. Such a tool would enhance instructor's control over the peer assessment process, while at the same time increase students' awareness and responsibility of their own learning. Therefore, in this article, we present an innovative, fully fledged peer assessment system, LearnEval, which aims to provide a solution to the existing gaps in the current landscape of systems, providing various affordances for both teachers and students. An initial version of the platform was introduced in [1] and [2], together with some pilot exploratory studies reported in [3], [5], and [6]. The current article presents the extended version of the platform, based on an iterative development process informed by the preliminary experimental findings. The platform validation is conducted in a project-

Manuscript received 24 April 2021; revised 7 March 2022; accepted 14 June 2022. Date of publication 22 June 2022; date of current version 5 August 2022. This work was supported by the grant POCU380/6/13/123990, co-financed by the European Social Fund within the Sectorial Operational Program Human Capital 2014–2020. *(Corresponding author: Elvira Popescu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Craiova and performed in line with the institution's ethical research guidelines.

The authors are with the Computers and Information Technology Department, University of Craiova, 200585 Craiova, Romania (e-mail: gabriel.badea@edu.ucv.ro; elvira.popescu@edu.ucv.ro).

Digital Object Identifier 10.1109/TLT.2022.3185423

based learning (PBL) scenario. In PBL, students learn by working individually or in groups to solve challenging tasks related to complex projects [38]. Joint application of PBL and peer assessment can further increase students' evaluation abilities and critical thinking skills [11], [19].

The current article adds to the literature by providing a comprehensive analysis of the peer assessment process supported by LearnEval. The evaluation takes into account three main directions: 1) validity of the grading process; 2) usability of the platform; and 3) subjective student satisfaction with the peer assessment experience. More specifically, we address the following research questions.

1) How did students participate in the peer assessment enhanced PBL scenario?
2) What are the validity levels of the peer grading process?
3) How did students perceive their experience of using LearnEval platform, in terms of usability and satisfaction?
4) What are the challenges faced by the students with the peer assessment approach?
5) How does students' experience with the extended version of the platform compare with the pilot studies?

The article continues with an overview of related work in Section II, providing a summary of existing peer assessment systems and highlighting their affordances and limitations. Section III describes the LearnEval platform, including a rationale, main functionalities, and implementation details, followed by an illustration of a generic peer assessment workflow. Next, Section IV presents a case study on the practical application of the platform in the context of PBL, addressing issues such as: peer assessment process validity, system usability, students' satisfaction with the learning experience, and a comparison with a similar scenario from the previous year. Section V outlines some conclusions and provides directions for future work.

II. Related Work

In what follows, we present an overview of the most representative peer assessment systems developed over the last two decades, highlighting important affordances and shortages they exhibit. The landscape includes two main categories of platforms: 1) systems aimed at assessing teammates in team-based assignments; and 2) systems aimed at assessing other peers in the class. Some of these platforms are more generic and can be applied to a wide range of pedagogical scenarios and contexts, while others are purposefully designed for a specific niche or discipline, such as programming, communication, or writing assignments.

A. Team-Based Peer Assessment

SPARK [15] is one of the earliest web-based platforms designed for peer assessment (and self-assessment) of student teamwork that attempts to solve the problem of free-riders. Self-assessment and peer assessment are applied by each member to polish the team's grade according to the individual contribution. Instructor's time is saved by automating tasks such as the collection of students' ratings or grade computation considering weight factors. Several settings can be configured by the teacher: assessment deadlines, review criteria, or team creation. Two types of assessment criteria are supported by the system: multiple criteria and holistic rating scheme. Furthermore, a bank of items can be reused by the instructor when defining the assessment criteria. At the end of the process, various metrics are computed. Help information and frequently asked questions (FAQs) are available for learners in case they need fast assistance. The system also allows learners to practice review. Some of the challenges at the time were the low external web access and the usability issues of the system.

A peer assessment platform similarly used for assigning grades that indicate the actual contribution of a peer in group work is SPAS [33]. The system features three types of users: administrator, teacher, and student. The admin is in charge of registering the users and handling tasks related to user management. The teacher creates the course and groups, defines the review criteria, and grades the learners. The student evaluates the quality of the work of other group members. The teacher might consider the peer review data and assign individual student grades based on it. At the end of the process, each learner has access to their individual final grade. A satisfaction questionnaire applied to the students revealed that they benefited from more peer collaboration and they were satisfied with the approach.

Another online peer-moderated marking tool specifically designed to facilitate group work is WebPA [24]. Just like in SPARK system, the students self-review and peer review their team members, "weighting factors" being generated which influence the group marks. The system can be applied to any type of group assignment. The instructor can set a broad range of parameters, such as: group size, number of groups, review criteria, or deadlines. Considering the high number of settings that need to be configured, new teachers can choose to apply default settings. At the end, the instructor provides grades to the teams and WebPA further produces reports describing the score assigned to each individual student. An important challenge is when a learner withdraws from a course affecting the performance of the group and the peer-moderated marking. Therefore, the instructor must manually inspect the results in groups where the members did not provide the evaluations.

Finally, Comprehensive Assessment of Team Member Effectiveness (CATME) is a set of three web-based tools that can be applied to assess achievement of learning goals related to teamwork [25]. Team-Maker tool gathers information about students and allows instructors to create teams. Past data can be employed to make more informed decisions about the criteria to apply for defining the teams. Peer Evaluation is a tool used for self-assessment and peer assessment of teammates' contributions. The system offers various statistics for the instructor computed based on the student assigned ratings. The instructors are automatically notified in case special conditions that might require further examination are met, such as ratings suggesting team conflict. The last tool, Rater Calibration, allows students to practice rating fictitious teammates and getting accustomed with the rating criteria before evaluating their actual teammates.

While the previous four systems are generic and can be used for any kind of team-based work, domain-specific platforms

have also been proposed in the literature. For example, a peer assessment system dedicated to medical students, aimed to solve the free-riders problem, is presented in [8]. The tool was employed to study whether the teams would benefit from higher contribution and cooperation. Two essential features were the open feedback and the balanced peer marking (i.e., the process where the students must come to an agreement and achieve consensus in assigning a grade to one of the team members). Each student holds a fixed amount of points to distribute and they are required to assign a different number of points to each teammate according to their contribution. The findings disclose that there were no differences in terms of team contribution and functioning between the learners that used the tool and the rest. Furthermore, focus group discussions revealed students' negative feelings regarding the process and the fear that the open, transparent evaluation might be considered threatening by the peers.

As can be seen, the main goal of team-based peer assessment platforms is to help the teacher clearly determine the contribution of each member to the teamwork, based on the teammates' appraisal; this also helps to eliminate the "free riding" problem. The underlying idea is that students are aware of their teammates' contribution, as they closely collaborated for the group assignment. The focus of this article, however, is on the more general type of peer assessment, in which each student is asked to evaluate the independent work of one or more peers. The underlying idea is that all students are supposed to solve a particular assignment and then be able to evaluate each other's solutions. Several examples of such platforms are surveyed in the next subsection.

B. Individual Peer Assessment

A generic, configurable, and powerful web-based self-assessment and peer assessment platform (called Web-SPA) is presented in [35]. Some of the main functionalities provided by the system are: account creation, assignment upload, scores and comments, progress tracking, and results visualization. The instructor holds a high degree of control by arranging the activities based on the objectives of the course or by monitoring students' reviewing activities. Various parameters can be configured, such as: assignment settings, review criteria, whether the process should be anonymous, type of marking, or group management. An experimental study showed a significant consistency between the grades assigned by the students through peer assessment and the instructor's scores. Moreover, the quality of the learners' work was enhanced as a result of the peer evaluation process.

A different approach, specifically designed for complex writing assignments, is proposed in [40]. The article describes a guidance system that can be incorporated in any peer assessment platform and aids evaluators to amend their reviewing skills and enhance the proficiency of their assessments. The work under review is automatically scaffolded such that the important content is drawn to the attention of the assessor. The instructor can set the review criteria and scaffolding regulations that are utilized by the tool. The innovation is provided by the mix between the algorithms employed to automatically locate the content of interest that needs the evaluator's focus and the final review decision that needs to be taken by the assessor. The outcomes of the experimental study aim to show whether the grades assigned by the peers are similar with the scores provided by expert evaluators.

Another interesting platform, called web-based formative peer assessment system (WFPAS) is presented in [20]; its main aim is to raise learners' metacognitive awareness and accomplishments in ill-structured tasks. The platform features rubrics for resolving ill-structured instructional design tasks and instruments for offering peer feedback. A unique feature represents reciprocal feedback, the possibility for reviewers and reviewees to exchange feedback in real time. The students must revise their initial draft based on the received feedback. Furthermore, they are required to offer feedback to the received reviews. In a final reflection phase, learners meditate on their own work and learning. The experimental findings reveal that the students who applied WFPAS obtained higher levels for metacognitive awareness, achievement in ill-structured tasks, and motivation, compared with the learners who participated in a traditional peer assessment process.

A system specifically built for assessing online videos is proposed in [22]. The students first upload the videos on an external website, a YouTube channel, and then they add the link in the peer assessment module of the system. The evaluators can attach comments and feedback to various moments in the video which subsequently can be visualized by the authors. The system was used in practice to assess the communication abilities of nursing students. Results showed a high level of correlation between the peer review grades and expert assigned scores. A questionnaire applied to the learners revealed that they were satisfied with the peer assessment method and appreciated its role in enhancing their communication skills. Furthermore, the students regarded peer assessments as fair and objective.

A peer assessment system dedicated to programming language learning is introduced in [41]. Educational Peer Code Review (EduPCR), as it is called, supports multipeer assessment, the procedure in which multiple evaluators review the same work. Some of the most important features of the platform are: automatic reviewer assignment mechanism, automatic score calculation, deadline tracking, revision, back-review, various scores, and automatic notifications. The weight of each score can be configured by the instructor according to the aims of the course. The system attempts to alleviate a part of the negative effects of reviewing by automatically detecting nonconsensus between the assessors and the radical behavior exhibited by some of the reviewers. The responses from a student questionnaire and an interview revealed that participants considered receiving multiple evaluations more useful than a single one and the approach can significantly enhance student learning achievements.

Another system dedicated to computer skills training is described in [17]. The approach requires students to build their own assessment forms in order to improve their higher order thinking skills. The peer assessment process was carried out anonymously. The findings of the experimental study revealed that the learning achievements of the students using the tool were significantly higher compared to the students who applied a conventional peer assessment system or no peer assessment at all.

In addition, a peer assessment platform where students review concept maps and are able to back-review or answer to evaluators is proposed in [18]. The bidirectional feedback increased the interaction between peers and the exchange of ideas. The experimental findings uncover that the system enhanced students' learning achievements and critical thinking, as well as the quality of the feedback they provide.

Finally, there is also a commercial web platform, called Eli Review, which was designed for scaffolding peer feedback activities, being particularly suited for online writing instruction [21]. The feedback strategy supported by the platform is grounded in writing pedagogy and consists of three phases: perform small writing assignments, attend peer assessment activities, and create revision plans based on the received feedback. The instructors focus on creating the tasks (writing tasks, review tasks, and revision tasks), set up debriefing sessions with the class about the peer feedback, and offer feedback to reviewers. Eli Review allows students to rate the quality of the received feedback and persuades instructors to provide comments on students' feedback. Qualitative and quantitative data are collected and made available to instructors through various reports, both in real time and for offline use.

C. Summary and Challenges

In addition to the classification presented above, the literature also distinguishes between peer assessment of product and peer assessment of process [31]. Peer assessment of product refers to students' evaluation of the quality of peers' work (output) using predefined criteria and assessment guidelines. By contrast, peer assessment of process refers to students' evaluation of the behavioral aspects of group work; this includes assessing their peers' teamwork skills and interaction, such as contribution to discussion and ideas, participation in group meetings, or feedback provision. Thus, peer assessment of a product is frequently applied in both individual- and team-based peer assessment systems; however, peer assessment of process is specifically tailored for team-based contexts and can provide relevant insights to the instructor for assigning grades that accurately appraise the actual contribution of each team member.

To sum up, a variety of peer assessment platforms have been proposed so far in the literature, being applied in various areas, such as computer science, medicine, literature, or physics. Many of these systems were designed to support a specific type of learning activity (e.g., complex writing assessments), a specific discipline (e.g., programming languages), or a specific type of student artifact (e.g., online videos). Furthermore, most of the platforms focus on a limited set of functionalities, in accordance with their specific goal. Hence, there is a need for a more comprehensive and generic platform, which should integrate useful functionalities present in various existing systems, but also novel desirable features.

Thus, important functionalities scarcely found in current platforms include: highly configurable peer assessment workflow, instructor's option to choose from several automatic and manual review allocation mechanisms, student modeling capabilities, and student's right to request an alternative expert opinion. Calibration/training options and anonymity of the reviewers and reviewees can play an essential role in lowering the pressure on the evaluators. The literature generally recommends that peer assessment process be double blind as this has several advantages [27]; however, a system should allow this setting to be configured based on the instructor's decision. In addition, automatic notification messages could prove very useful, drawing users' attention when important events occur in the platform. Finally, various intuitive statistics, scores, and comprehensive open learner model (OLM) visualizations are highly desirable features, as they allow both the student and the instructor to easily track the learning progress. Starting from this literature analysis and needs assessment, we designed and implemented a comprehensive and generic peer evaluation system that accommodates common functionalities found in existing platforms, but also additional features that enhance both the teacher and the student experience, as described in the next section.

III. LearnEval Peer Assessment Platform

A. Rationale

LearnEval is a peer assessment system conceived with the main aim to provide both learners and instructors a tailored and rich experience. The platform belongs to the *individual peer assessment category*, focusing on the *peer assessment of product*, according to the classifications mentioned in the previous section. LearnEval was designed and implemented as a solution to the existing challenges in the current peer assessment systems and to fulfill the requirements of every type of course. The first step in developing the platform was to devise a peer assessment workflow that would be suitable in a wide range of pedagogical scenarios and contexts but, at the same time, would be sufficiently configurable to satisfy the requirements of any particular scenario. The instructor can define various aspects related to the process, such as: deadlines, assessment criteria, number of reviews per submission, mechanism employed for allocating the submissions to reviewers, anonymity of the reviewers and reviewees, and weights of the various metrics used for computing scores and modeling learners [2]. Furthermore, the instructors can track the progress of the learners to detect students at risk, visualize learner models for assessment purposes, create calibration assignments for enhancing students' assessment skills, or receive notifications when actions of interest occur. On the other hand, the students' experience is enhanced and enriched by features, such as: back-review to assess the quality of the received reviews stimulating critical thinking, access to OLM, or intuitive statistics of the peer assessment data [1].

The initial version of LearnEval was applied in several pilot courses [3], [5], [6]. The data gathered from the platform, together with the students' answers to satisfaction and usability surveys, helped us improve LearnEval and, thus, enhance the experience of both instructors and learners. The findings of the pilot studies also provided guidelines for refining the peer assessment process in order to achieve higher levels of grading validity and learner satisfaction, as discussed in the next sections.

B. Initial Prototype

The first version of the LearnEval platform integrated core functionality required in a generic peer assessment system. In the following, we present the main modules and features of the teacher and student areas, respectively.

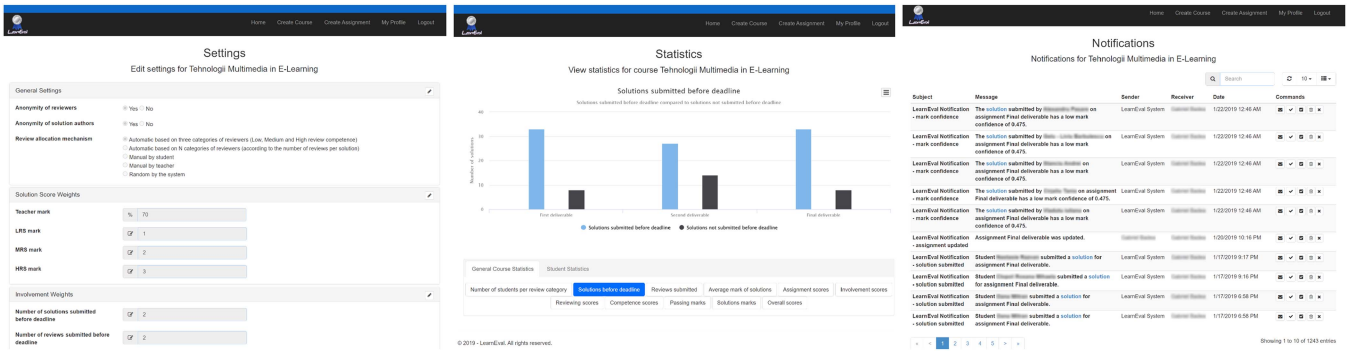

Fig. 1. Screenshot collage illustrating some of the teacher functionalities in LearnEval (course settings, statistics, and notifications).

LearnEval offers instructors a practical and versatile management of the peer assessment process. The teacher area is structured in modules such that every aspect can be readily accessed and managed [2].

1) Course

The course represents the framework in which the peer assessment process takes place; each unit is labeled with a name, year of study, and description. The instructor creates a course, enrolls students to it, and devises the peer assessment sessions.

2) Assignments

The instructor can create and manage assignments for the taught courses. An assignment has several properties, such as: type, name, requirements, submission and review deadlines, number of assessors per submission, and assessment criteria. This module also displays the solutions submitted by the students for each assignment, with the option for the instructor to evaluate them.

3) Settings

The peer assessment workflow can be configured based on various parameters, such as: anonymity of reviewers/reviewees (solution authors), mechanism to allocate the submissions to reviewers, and weights of the metrics used for computing student scores. Several mechanisms for allocating the submissions to reviewers are supported by the LearnEval platform, such as: automatically based on the assessment skills of the reviewers, randomly by the system, or manually by the instructor or learners. Default values are proposed for the various configuration parameters in order to provide some guidelines/help for the instructors. For example, instructor evaluation represents 70% of the final grade assigned to the submission by default; however, the instructor can set it to a lower or higher value. A weight of 0% lets students' reviews to represent the full grade, whereas a weight of 100% is applied in cases when peer assessment is designed only for formative purposes and not for final grading.

4) Reviews

This module provides access to the reviews submitted by students. These reviews can be filtered based on various criteria, such as: reviewer name, assignment name, review date, or review category. The instructor can provide a back-review to each student review, thus offering the necessary feedback.

5) Statistics

This module helps the instructor easily examine various aspects of the peer assessment activity data. Two types of statistics are available: *General Course Statistics* and *Student Statistics*. *General Course Statistics* section provides information regarding the class as a group: number of solutions submitted before deadline for each assignment, mean assigned grade for each assessment session, number of submissions with a passing grade, or student scores. *Student Statistics* section offers information regarding a particular selected student: grades assigned to the solutions submitted by the student, number of submitted reviews for each solution, or number of received back-reviews.

6) Scores

LearnEval models students based on several characteristics, such as: involvement, reviewing skills, and competence. Each characteristic is further decomposed in several relevant metrics. The *involvement score* depicts the level of student's participation in the peer assessment activities and depends on the number of submitted solutions, reviews, and back-reviews. The *reviewing score* portrays the student's assessment capabilities and relies on the accuracy of the offered evaluations and the subsequently received back-reviews. Finally, the *competence score* illustrates the proficiency with the topic and depends on the grades assigned to the student's submitted solutions. The instructor can access all these scores under two perspectives: a numerical and a graphical representation of the learner model. The provided model is very comprehensive, allowing the teacher to see at a glance the areas that need intervention.

7) Notifications

The system sends notifications when relevant actions occur, such as: a solution was submitted, a grade was assigned with a low confidence factor (which may require instructor's attention), or the submission/review deadline is approaching. A special type of notification is the request for an expert opinion; these are appeals from students to instructors to review their submissions in case they are not satisfied with the received peer evaluations.

Some of the above functionalities which LearnEval provides for the instructor are illustrated in Fig. 1.

LearnEval also empowers students by offering a rich peer assessment experience. The student area is carefully designed such that every aspect of the process can be easily accessed and examined [1].

1) Assignments

This module displays the available assignments. A solution can be submitted by the students to the assignments whose

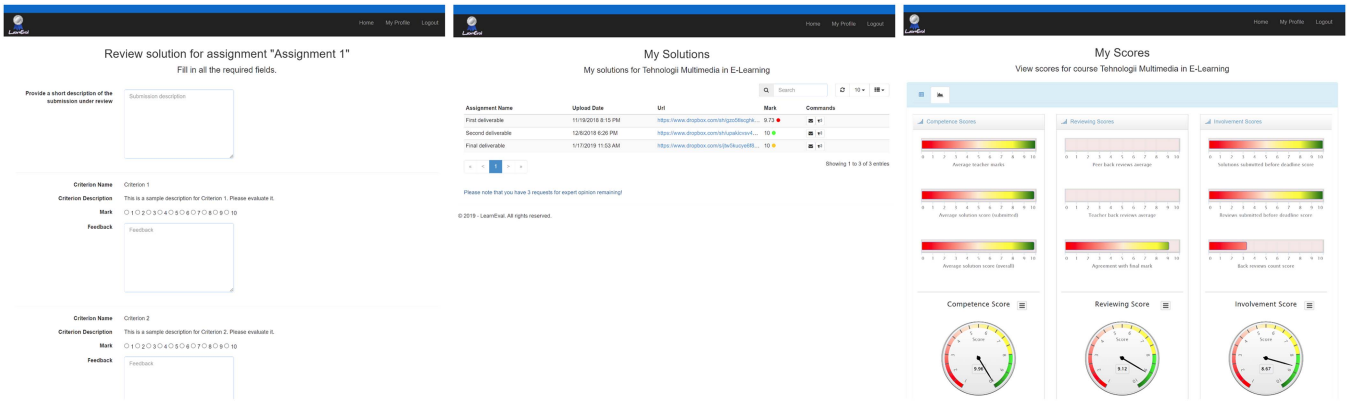

Fig. 2. Screenshot collage illustrating some of the student functionalities in LearnEval (review solutions, view solutions, and scores).

deadline is still open; more specifically, a URL needs to be provided, from where the solution can be downloaded by the reviewers.

2) Review Solutions

In this module, the student has access to the allocated submissions to review. The student evaluates each submission and fills in an assessment form that includes several assessment criteria defined by the instructor; each criterion consists in a grade on a scale from 1 to 10, plus a textual feedback. Furthermore, the student must also provide a short description (overall evaluation) of the submission under review.

3) My Solutions

This module shows the solutions submitted by the student, together with their assigned grades, once they are available. The solution grade is automatically computed by the system, as a weighted sum of the peers' and teacher's grades, taking into account the reviewing skill levels of the students; the actual weights can be configured by the teacher in the course settings. A confidence factor is also computed for each grade, indicating the level of trust for that grade, based on the reviewing skills of the evaluator students. A link to the received reviews is also available for each submission. Furthermore, the student can provide a back-review for each received review, assessing its quality, helpfulness, and appropriateness. The back-review includes a grade on a scale from 1 to 10, plus a textual feedback to help evaluators reflect on the impact of their reviews.

In the special case when a student is not satisfied with the peer evaluations received, requests for expert opinion (i.e., teacher evaluation) can be sent from this module. The number of such requests is limited to three, in order to avoid misusing this feature. Furthermore, the learner must offer a clear reason associated with each request. However, in case the teacher considers the appeal appropriate, the student is rewarded with one extra expert opinion request.

4) My Reviews

This module displays the reviews submitted by the student, together with the received back-reviews. Furthermore, the student has access to the final grade assigned to the submission they evaluated, together with the reviews submitted by other peers for the same solution; this provides opportunities for comparison, stimulating curiosity, and encouraging critical thinking.

5) Statistics

This module offers the same information available in the *Student Statistics* section from the teacher area: grades assigned to the solutions submitted by the current student, number of submitted reviews for each solution, or number of received back-reviews.

6) Scores

Again, the information available in this module is similar to the one from the instructor area (for the current student). An OLM is provided, which gives the student access to an intuitive and effective visualization of their own activity and scores. OLMs are learning awareness tools that offer students real-time information regarding their learning; they make use of a machine representation of the learner to reveal student's progress and achievement [10]. In LearnEval, the learner is modeled based on involvement, competence, and reviewing scores, which are also aggregated in an overall score. The student can visualize their OLM both in textual format (which presents the raw numerical values obtained by the learner) and in graphical format (which displays the model using various visual components, such as progress bars, column bars, gauges, trophies, and medals) [4].

7) Notifications

This module sends notifications to the students when relevant actions occur, such as: a new assignment is open, the submission/review deadline for an assignment approaches, or a grade was assigned to one of the students' submissions. The notification messages contain direct links to the corresponding sections of the system, for a straightforward access.

Some of the above functionalities which LearnEval provides for the student are illustrated in Fig. 2.

C. Pilot Studies and Platform Extensions

In the 2018−2019 academic year, LearnEval was applied in the context of three pilot courses at the University of Craiova, Romania; this allowed us to collect both usage data from the platform and students' subjective opinions from dedicated surveys. In the first semester, LearnEval was used in a Multimedia Technologies in E-Learning course [3], while in the second semester it was used in a Web Applications Design course [5] and a Human−Computer Interaction course [6], with some variations between the applied scenarios. Based on the findings from the three studies and the feedback provided

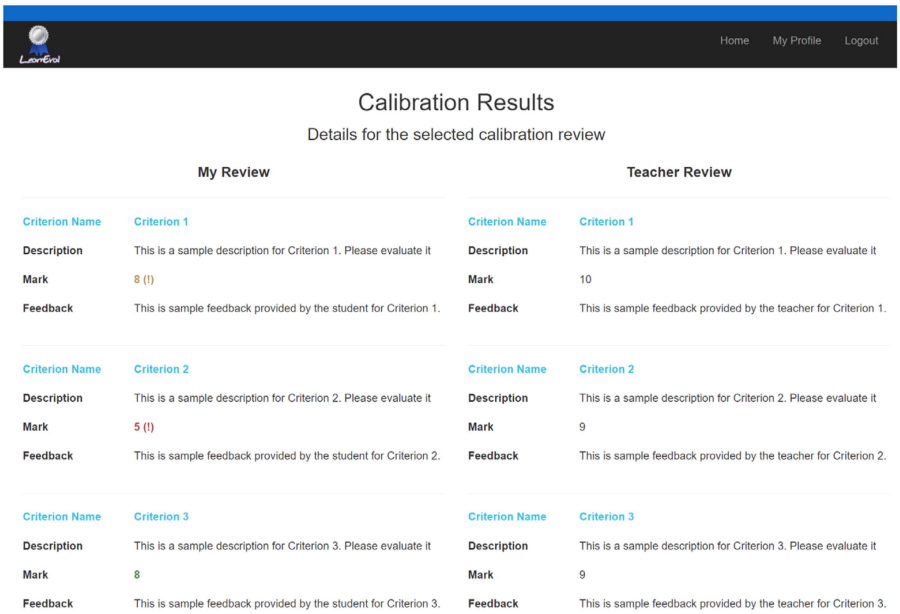

Fig. 3. LearnEval sample calibration results page.

by the students and instructors, the platform went through an extension and improvement process; the following main changes were performed in the second version of LearnEval.

1) A calibration module was developed in order to allow students to practice and enhance their assessment skills before the start of the actual reviewing process.
2) A HowTo module comprising FAQs was integrated in order to increase the friendliness of the platform and to provide a straightforward way to find different features.
3) A more efficient approach for the visualization of the reviews by the teacher was designed.
4) The minimum required feedback length was increased from 10 to 50 characters in order to encourage reviewers to offer more comprehensive and helpful feedback.
5) Several small bugs and technical issues were solved.

In what follows, we present more details regarding the most important extension of the platform, namely the *calibration module*. Calibration or training allows learners to gain experience with reviewing and enhance their skills [42], view samples of high-quality work, and increase their confidence to evaluate others [32]. Calibration can solve problems related to assessment quality, such as grading bias or rogue reviews [32]. The activity offers students the chance to practice, test, and improve their assessment skills before reviewing actual peers' work. The students get accustomed with the review rubric and offer more accurate grades, reducing instructor grading workload. Furthermore, the instructor might restrict the participation to the actual peer assessment sessions for the students not demonstrating the required reviewing skills during training. Several calibration tools have been reported in the literature: Rater Calibration [25] provides students the opportunity to rate fictitious teammates and get used to the review criteria before assessing their actual peers. In Mechanical TA [42], students can practice reviewing by assessing essays from past offerings of the course and compare their evaluations with the "gold standard" reviews performed by teaching assistants. The reviewers become more independent by attending calibration sessions and their evaluations do not need to be checked by the teaching assistants; thus, the time dedicated by instructors for the assessment task is reduced.

The proposed LearnEval calibration module allows students to assess dedicated test submissions and compare their reviews with reference (model) reviews provided by the instructor. Learners can, thus, analyze the similarities and differences between their reviews and the ones written by the teacher; this leads to a better understanding of the expectations of the peer assessment process and promotes critical thinking skills. Fig. 3 illustrates the comparative view of the calibration results, as shown to the students in LearnEval. As can be seen, the deviation between the grades assigned by the student and the ones assigned by the instructor is highlighted through colors: green (maximum 1 point difference), yellow (2−3 points difference), or red (more than 3 points difference). Furthermore, a *calibration score* is computed for each student, based on the closeness of the learner's grades and the ones provided by the instructor, for all the calibration assignments. This metric is accessible from the *scores* module and is taken into consideration when computing the *reviewing score* of the student, according to the weight configured by the teacher.

D. System Architecture and Implementation

From the start of the development process, LearnEval was designed to be robust and easy to extend and maintain. The system follows a classic three-layer architecture and is built around modules in order to support iterative development and evolving requirements. In terms of implementation, the system was developed in C# using ASP.NET MVC 5 framework on the server side, and JavaScript and KnockoutJs on the client side. The data is stored in an SQL Server Database and the access is performed using Entity Framework ORM. Dedicated

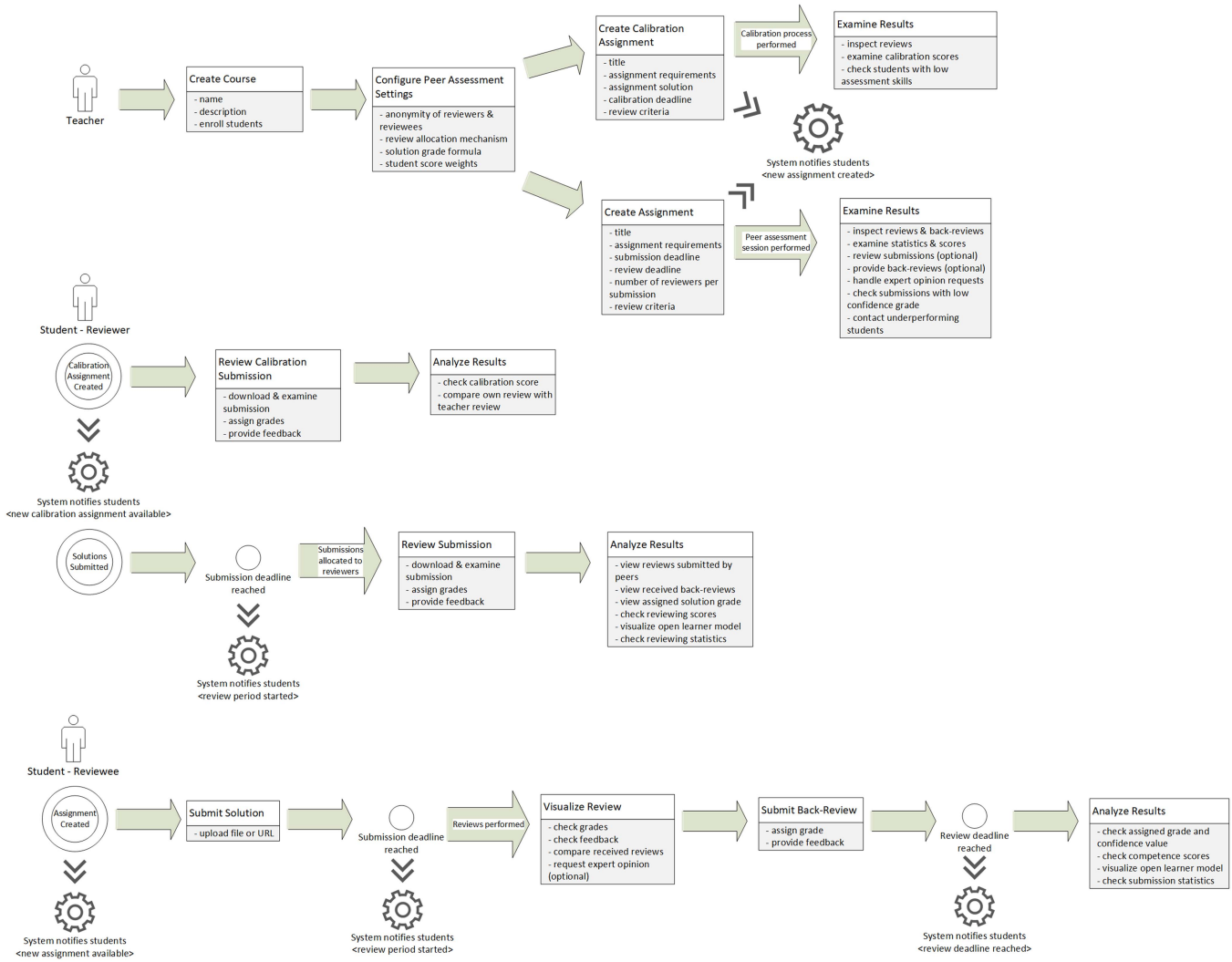

Fig. 4. Basic peer assessment workflow in LearnEval.

projects exist for each of the three core layers of the system: presentation, business, and data access. Furthermore, several other projects were developed for the recurring tasks. Various design patterns, such as unit of work or repository pattern, were employed for a better maintenance. Finally, the user interface is friendly and responsive, the platform being easily accessible from mobile devices.

E. Illustration of a Generic Workflow

As mentioned above, LearnEval is a versatile and generic platform, allowing instructors to apply it in a broad range of courses, contexts, and scenarios. In this section, we provide an example of how the system can be used to support a basic peer assessment process. The workflow is illustrated in Fig. 4 and consists in the following steps.

1) Course Setup: First of all, the instructor must define and set up the course in the platform.

a) The instructor creates the course by specifying its name, year of study, and a short description.
b) The instructor configures parameters related to the peer assessment process (or uses the default configuration settings): anonymity of the reviewers and reviewees, mechanism employed for allocating the submissions to reviewers, weights for computing the submission grade, and weights of the metrics used for computing student scores.
c) The instructor enrolls the students to the course.

2) Calibration/Training Phase:

a) The instructor creates the calibration assignment by providing reference reviews for model submissions. A notification is sent to each student that a calibration assignment has been created.
b) The students evaluate the training submissions and compare their reviews with the ones provided by the teacher.
c) Both the instructor and the students can analyze the outcome of the calibration stage.

3) Peer Assessment Sessions: The instructor generates the assignments that need to be delivered by the students and devises a peer assessment session for each one, as follows.

a) The instructor creates the assignment by specifying its name, requirements, submission and review deadlines,

and assessment criteria. A notification is sent to each student that a new assignment has been created.

b) The students have a submission period to provide solutions to the assignment. A reminder notification to submit the solution is sent to the students 24 hours before the deadline.
c) When the submission deadline is reached, each submitted solution is allocated to evaluators depending on the review allocation mechanism set by the instructor.
d) The students have a review period to assess peers' submissions. Likewise, a reminder notification to submit the reviews is sent to the students 24 hours before the review deadline. The submission author is notified whenever a review is received.
e) Each submission is assigned a grade and a confidence factor based on the received evaluations. A notification is sent to the instructor when a low confidence factor is recorded, so that they can examine the corresponding submission and reviews.
f) The instructor analyzes the various outcomes of the peer assessment process. The teacher can examine statistics and scores at course level or delve into details and visualize results related to a particular student. Supportive measures for specific learners could be planned based on this analysis.
g) The learners have access to the grades assigned to their submissions and they can provide back-reviews for the received evaluations. Furthermore, the student can access their OLM, examining personal scores and visualizing statistics. Therefore, areas that need improvement can be easily detected and corrective actions can be taken accordingly.

The process can be repeated for any number of assignments per course, each with its dedicated peer assessment session.

In the following section, we illustrate a practical application of LearnEval in a more complex PBL scenario and explore the results of the peer assessment process.

IV. Case Study: LearnEval and PBL

PBL is a student-centered learning technique aimed at producing meaningful learning experiences, where students take more responsibility and act more independently [14]. PBL is a pedagogical method frequently applied in higher education, incorporated into educational technology, and a popular research topic [16]. The approach is devised around projects and defined by the following characteristics: student focus, driving question, constructive investigation, autonomy, and realism [38]. PBL offers many benefits for learners such as increased motivation and responsibility, more independent learning, enhanced transversal and discipline skills, and enhanced critical thinking. Furthermore, learners gain a better understanding of the professional practice and how to employ obtained knowledge to real-world problems [29]. PBL uses a challenge to enable inquiry activities where learners address questions, look up information, and design and test various alternative solutions [9]. Students create complex artifacts by using what they have learned or found during the inquiry. These artifacts can be evaluated and discussed for further improvement [14].

The generic peer assessment workflow supported by LearnEval allows it to be easily employed in courses following a PBL scenario. The several project milestones throughout the semester, each with its own project deliverable, can be used to devise effective peer evaluation sessions. Such a peer assessment enhanced PBL scenario is presented in this section, aiming to provide an experimental validation of the LearnEval platform from three perspectives: peer grades validity, system usability, and subjective learner satisfaction.

A. Course Context

The extended version of LearnEval platform was applied in the Multimedia Technologies in E-Learning course during the first semester of the 2019−2020 academic year. The course is taught to fourth-year undergraduate students following a degree in Multimedia Systems Engineering at the University of Craiova, Romania. The course implied a PBL component requiring students to individually develop projects. The task of the project was to implement a website presenting an Informatics lesson for high-school level. The first milestone required students to create the structure of the website and the hierarchy of pages and present theoretical notions related to the chosen Informatics algorithm. In terms of technologies, the students had to make use of HTML for building the layout and Cascading Style Sheets (CSS) for styling the pages. The second milestone required learners to implement the chosen algorithm in JavaScript programming language and to integrate multimedia material related to it, such as images, audio, and video content. For the last milestone, the students had to develop educational games, interactive simulations, and a knowledge assessment test.

The instructor decided to use the three milestones to create three peer review sessions. The following peer assessment settings were employed: the process was double-blind; the review criteria were the same in all sessions and referred to functionality, implementation, and esthetic and pedagogical quality. The peer assessment activity was mandatory, representing 30% of the final grade assigned to the project. Each submission (i.e., project deliverable) was allocated to three reviewers with various assessment skills. This allocation is performed automatically by the system once the submission deadline defined by the instructor has passed. The reviewers are divided into three categories reflecting their assessment capabilities based on the *reviewing score*: students with high reviewing skills, students with medium reviewing skills, and students with low reviewing skills. One reviewer from each category is assigned to each submission, which leads to a fair distribution of the evaluations. In addition, the allocation mechanism selects from each category the student with the lowest number of reviews assigned across all sessions; hence, by the end of the course, all students are allocated a similar number of evaluations to perform (differing by at most one).

As an enhancement compared to the pilot studies conducted in the previous year, a calibration/training phase was included in the scenario, which took place before the first project assignment. The teacher prepared two reference project

TABLE I
NUMBERS AND PERCENTAGES OF PROJECT DELIVERABLES AND REVIEWS SUBMITTED FOR EACH ASSIGNMENT

Assignment	I	II	III	Total
Deliverables	18 (62%)	22 (76%)	21 (72%)	61
Reviews	47 (87%)	46 (70%)	40 (63%)	133

TABLE II
STATISTICS ON CENTRAL TENDENCY MEASURES AND CENTRAL DISTRIBUTION OF THE GRADES ASSIGNED BY THE STUDENTS VS. TEACHER

	Student grades	Teacher grades
Mean	8.25	7.81
Median	8.53	8
Standard deviation	1.46	1.75

deliverables for students to practice evaluation on: one with high quality and one with low quality. The goal was to help students identify both strong and weak points and provide adequate feedback in each case. Approximately half of the students attended this calibration phase and reviewed both practice deliverables. The difference between the learner assigned grades and the reference grades provided by the teacher ranged between 0 and 3, indicating different initial reviewing skills of the students.

B. Participation Results

This section aims to address our first research question: "How did students participate in the peer assessment enhanced PBL scenario?" Out of the 30 students enrolled in the course, 29 registered in the LearnEval platform (i.e., 97%[1]); 27 of them attended at least one of the peer assessment sessions (i.e., 93%).

Table I depicts the number of artifacts (project deliverables and reviews) uploaded by the students. Interestingly, while the number of project deliverables was higher for the last two sessions, the number of reviews was lower.

In terms of grading, in general, students were more lenient than the teacher by assigning higher and more narrowly dispersed grades, as shown in Table II. The column "Student grades" refers to the weighted grade computed by LearnEval, based on the individual peer grades.

As far as the textual feedback is concerned, its length is relatively short, as shown in Table III. Out of the total of 399 comments, only 59 (i.e., around 15%) exceed 200 characters and less than 5% exceed 400 characters.

In what follows, we provide a few examples to illustrate the different quality levels of the feedback given by the students. On one hand, there are some short and not very relevant comments, which are not particularly helpful for the solution authors, such as: "The site works properly, the buttons and menu work," "It is correctly implemented. The points related to JavaScript, HTML and CSS have been reached," "It looks good from an aesthetic point of view and contains a lot of information," or "It contains all the features that were required and even more. The site is responsive." On the other hand, there are also more elaborate and useful comments that

[1] Percentages are rounded to the nearest integer throughout the article.

TABLE III
DISTRIBUTION OF THE LENGTH OF STUDENT FEEDBACK

Length (number of characters)	< 100	[100 - 200)	[200 - 400)	>= 400
Comments count	250 (63%)	90 (23%)	40 (10%)	19 (5%)

TABLE IV
CORRELATION BETWEEN PEER GRADES AND TEACHER GRADES

Assignment	I	II	III	Overall
Correlation Coefficient	0.92	0.79	0.51	0.72

emphasize the strong and weak points of the submission and offer suggestions for improvement, such as: "The site is of the single page type with very little information and presented in an incomplete manner. Only the Quicksort algorithm is included, which performs sorting based on a Divide et Impera strategy. The site does not provide a detailed definition of the Quicksort algorithm. I would have added more pages and more information for a more consistent overall structure. For example, I would have mentioned by whom the algorithm was created or I would have presented a possible implementation of the algorithm, in the C language in the form of a function. The menu is not functional, it does not direct us to any page," or "Tags are used accordingly. There is only one HTML page in which there is a separation between the HTML code and the CSS code, an external style sheet being used. There is no dedicated folder for resources, i.e., the image used is in the same folder as the HTML page."

C. Peer Grading Validity

Validity represents the level of agreement between the grades given by the students and the reference ones, given by the teacher. Attaining high levels of validity is essential for an effective peer assessment process. In order to answer our second research question ("What are the validity levels of the peer grading process?"), we computed the Pearson Correlation Coefficient between the grades assigned by LearnEval based on the peer reviews and the grades given by the instructor. Pearson correlation has been employed as a sound measure of the relationship between the two variables [23] and the instructor grades have been reported in the literature as valid expert reference metrics [28].

The results are included in Table IV. The overall correlation value, which considers all the peer assessment sessions, was quite high (0.72). The correlation was even higher for the first two milestones; the lower figure obtained for the last assignment was due to the existence of several rogue reviews that were assigning high grades to low quality deliverables, without offering any appropriate justifications.

D. System Usability

In order to assess students' perceived experience of using LearnEval (research question 3), we started with gauging their opinion on the usability of the platform. To this end, we employed the system usability scale (SUS) survey [12], which offers a reliable, easy-to-apply instrument for assessing the

TABLE V
RESULTS OF SUS SURVEY

SUS Item	*Percentage of responses*
1. I think that I would like to use this system frequently *(relative and strong agreement)*	52%
2. I found the system unnecessarily complex *(relative and strong disagreement)*	70%
3. I thought the system was easy to use *(relative and strong agreement)*	74%
4. I think that I would need the support of a technical person to be able to use this system *(relative and strong disagreement)*	61%
5. I found the various functions in this system were well integrated *(relative and strong agreement)*	70%
6. I thought there was too much inconsistency in this system *(relative and strong disagreement)*	61%
7. I would imagine that most people would learn to use this system very quickly *(relative and strong agreement)*	78%
8. I found the system very cumbersome to use *(relative and strong disagreement)*	74%
9. I felt very confident using the system *(relative and strong agreement)*	57%
10. I needed to learn a lot of things before I could get going with this system *(relative and strong disagreement)*	74%

usability of a system; it consists of 10 five-level Likert items, ranging from strongly disagree to strongly agree.

Twenty-three students (i.e., 79%) filled in the SUS survey at the end of the semester and the results are summarized in Table V. Odd rows display the percentages of students that selected either relative or strong agreement (as those items are worded in a positive way), whereas even rows display the percentages of students that selected either relative or strong disagreement (as those items are worded in a negative way). As can be seen, all figures are over 50% and many of them are at least 70% (items 2, 3, 5, 7, 8, and 10). The overall SUS score, computed according to [12], is 72.83, which qualifies the LearnEval usability as "good" according to [7].

E. Students' Satisfaction and Learning Experience

In order to further explore students' experience with the LearnEval platform and the peer assessment process (research questions 3 and 4), we designed and applied a dedicated satisfaction questionnaire. Twenty-one of the students (i.e., 72%) filled in this questionnaire at the end of the semester. In what follows, we summarize the main results obtained.

1) About Peer Assessment in General: The results show that students are generally open to the peer assessment process. Most of the learners consider it a good or very good idea to review peers' projects (71%) and to be reviewed by the peers (71%). Furthermore, most learners deemed helpful or very helpful to offer feedback to peers (71%) and to receive feedback from the peers (67%).

2) About the Settings of the Peer Assessment Activity: Most students appreciated the amount of time available for submitting the solutions (81%), reviewing peers' work (86%), as well as the number of assessment criteria (100%). Furthermore, the majority of the learners agreed that both the reviewer and the reviewee must be anonymous (67%). Finally, almost all students regarded the number of calibration submissions as suitable (91%).

3) About the Experience as Reviewer: More than half of the students liked to see the peers' submissions (52%) and to review them (57%) to a high or very high degree. Furthermore, a similar percentage of students agreed that back-reviews helped them to strengthen their reviewing skills (52%). Almost all students were neutral or liked the calibration module to a high or very high degree (95%). Moreover, most learners were neutral or considered that calibration reviews helped them to strengthen their reviewing skills to a high or very high degree (90%).

4) About the Experience as Reviewee: More than half of the students considered that the received reviews were objective and unbiased (62%), or complete and detailed (52%) to a high or very high degree. Furthermore, 67% of the learners agreed that the received reviews helped them to improve their projects.

5) About Motivation and Time: Around half of the students considered the peer assessment process motivating to a high or very high degree (48%). Only a small part of the learners deemed the process stressful (10%) or time-consuming (19%).

6) About Overall Experience: Overall, most of the students were satisfied or very satisfied with the LearnEval platform (71%) and would like to use the system in future courses (62%).

7) About Encountered Challenges: While learners were generally satisfied with the platform and peer assessment process, a few negative issues were also mentioned, e.g.,: "Subjectivism," "I do not like to assign grades, it would have been better if I only had to provide feedback," "The assessment could be unreliable," and "Most of the students do not have the required knowledge to assess a project."

F. Comparison With Pilot Study

Finally, we were interested to see how students' experience with the extended version of the LearnEval platform compares with the earlier pilot study (research question 5). As mentioned in Section III, the initial version of LearnEval was used in a previous edition of the Multimedia Technologies in E-Learning course (2018−2019 academic year) [3]. The scenario settings were similar between the two course runs. Several improvements could be noticed in the second year, as described next.

The validity of the grading process was significantly enhanced; the correlation coefficient increased from 0.19 in the first year to 0.72 in the second year. This could be due to the inclusion of the calibration module, which helped students practice their assessment skills before the start of the actual reviewing process. Furthermore, a detailed introductory session was integrated at the beginning of the semester, when the instructor explained the peer assessment process and its expected benefits.

The average length of student feedback was higher, which could be due to the increase in the minimum number of characters required in the corresponding form. The quality of the provided comments seemed to be higher as well, but there is still room for improvement.

The SUS score increased from 62.76 ("OK" category) to 72.83 ("Good" category); the lower figure in the first year was

caused by the small bugs and technical issues encountered, which were solved in the improved version of LearnEval. The addition of the *HowTo* module also likely contributed to a better understanding of the platform features.

Students' satisfaction with LearnEval and the peer assessment process also increased. For example, the percentage of students who considered it helpful or very helpful to offer feedback to peers more than doubled (34% in the first year vs. 71% in the second year); similar figures were reported for receiving feedback (34% vs. 67%). Back-reviews were also considered more helpful for strengthening students' reviewing skills (19% vs. 52%). Furthermore, a higher percentage of students considered that the received reviews were objective and unbiased (34% vs. 62%) or complete and detailed (21% vs. 52%) to a high or very high degree. In addition, more students agreed that the received reviews helped them to improve their projects (34% vs. 67%). Finally, the general comments provided by the students were more positive in the second year, compared to the first one, indicating a higher level of satisfaction with the learning experience.

V. Conclusion

This article covered the iterative development and successful experimental validation of a comprehensive and generic peer assessment platform. LearnEval was built as an alternative to the special-purpose, context-specific peer assessment systems reported in the literature, being highly flexible and able to accommodate a variety of pedagogical scenarios. The platform integrates a wide range of functionalities, such as: highly configurable peer assessment workflow, several automatic and manual review allocation mechanisms, initial calibration/training options for the students, back-reviews and expert opinion requests, complex dashboard with statistics, scores and graphical visualizations, as well as an extensive OLM.

The platform was used in a real-world PBL scenario, in a Multimedia Technologies in E-Learning course. The peer assessment process was investigated and reported in detail, from several perspectives: student involvement and activity levels, peer grades validity, perceived learning experience, and challenges encountered. Results are very encouraging and indicate a clear improvement compared with the pilot studies performed with an earlier version of the platform, in terms of system usability, grading validity, and student satisfaction.

The main limitation of our study consists in the relatively small number of participants; this is a convenience sample, based on the students enrolled in the Multimedia Technologies in E-Learning course in that particular academic year. In the future, we plan to use the platform in several other courses, with different pedagogical scenarios and disciplines of study, in order to perform a more comprehensive analysis of the peer assessment process.

Moreover, the quality of the peer feedback provided by the students could be further improved. In addition to continuous instructor scaffolding and providing ample training opportunities for reviewing, some automatic mechanisms could also be integrated in the platform. Thus, we aim to devise an approach for automatically detecting low quality and rogue reviews, to be implemented in LearnEval.

Finally, the current review allocation mechanism is static, i.e., the allocation is performed only once, at the beginning of the peer assessment session. This can lead to submissions receiving an unequal number of reviews, in the event that some students do not complete their peer evaluations. Therefore, a dynamic review allocation approach, which would provide a more balanced distribution of reviews, is another future research direction we envision.

References

[1] G. Badea and E. Popescu, "A web-based platform for peer assessment in technology enhanced learning: Student module prototype," in *Proc. IEEE 19th Int. Conf. Adv. Learn. Technol.*, 2019, pp. 372–374, doi: 10.1109/ICALT.2019.00115.

[2] G. Badea and E. Popescu, "Instructor support module in a web-based peer assessment platform," in *Proc. 23rd Int. Conf. Syst. Theory, Control Comput.*, 2019, pp. 691–696, doi: 10.1109/ICSTCC.2019.8885687.

[3] G. Badea and E. Popescu, "Using LearnEval peer assessment platform in project-based learning settings: A first experience report," in *Proc. 18th Int. Conf. Inf. Technol. Based Higher Educ. Training*, 2019, pp. 1–5, doi: 10.1109/ITHET46829.2019.8937359.

[4] G. Badea and E. Popescu, "Supporting students by integrating an open learner model in a peer assessment platform," in *Proc. Int. Conf. Intell. Tutoring Syst.*, 2020, pp. 105–114, doi: 10.1007/978-3-030-49663-0_14.

[5] G. Badea and E. Popescu, "Using peer assessment in conjunction with project-based learning: A comparative study," in *Proc. IEEE 20th Int. Conf. Adv. Learn. Technol.*, 2020, pp. 376–380, doi: 10.1109/ICALT49669.2020.00119.

[6] G. Badea and E. Popescu, "Analyzing the validity of the peer assessment process in a project-based learning scenario: Preliminary results," in *Proc. 24th Int. Conf. Syst. Theory, Control Comput.*, 2020, pp. 831–836, doi: 10.1109/ICSTCC50638.2020.9259671.

[7] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *Int. J. Hum.-Comput. Interaction*, vol. 24, no. 6, pp. 574–594, Aug. 2008, doi: 10.1080/10447310802205776.

[8] R. Biesma, M. C. Kennedy, T. Pawlikowska, R. Brugha, R. Conroy, and F. Doyle, "Peer assessment to improve medical student's contributions to team-based projects: Randomised controlled trial and qualitative follow-up," *BMC Med. Educ.*, vol. 19, no. 1, pp. 371–378, Oct. 2019, doi: 10.1186/s12909-019-1783-8.

[9] P. Blumenfeld, E. Soloway, R. Marx, J. Krajcik, M. Guzdial, and A. Palincsar, "Motivating project-based learning: Sustaining the doing, supporting the learning," *Educ. Psychol.*, vol. 26, no. 3/4, pp. 369–398, Nov. 1991, doi: 10.1080/00461520.1991.9653139.

[10] R. Bodily *et al.*, "Open learner models and learning analytics dashboards: A systematic review," in *Proc. 8th Int. Conf. Learn. Analytics Knowl.*, 2018, pp. 41–50, doi: 10.1145/3170358.3170409.

[11] M. Boubouka and K. Papanikolaou, "Alternative assessment methods in technology enhanced project-based learning," *Int. J. Learn. Technol.*, vol. 8, no. 3, pp. 263–296, Sep. 2013, doi: 10.1504/IJLT.2013.057063.

[12] J. Brooke, "SUS: A retrospective," *J. Usability Stud.*, vol. 8, no. 2, pp. 29–40, Jan. 2013, doi: 10.5555/2817912.2817913.

[13] H.-C. Chen, G. Pisoni, and H. Gijlers, "An exploratory study into graduate students' attitudes towards peer assessment," in *Proc. Int. Conf. Interactive Collaborative Learn.*, 2020, pp. 310–325, doi: 10.1007/978-3-030-68201-9_32.

[14] Y.-H. Ching and Y.-C. Hsu, "Peer feedback to facilitate project-based learning in an online environment," *Int. Rev. Res. Open Distance Learn.*, vol. 14, no. 5, pp. 258–276, Oct. 2013, doi: 10.19173/irrodl.v14i5.1524.

[15] M. Freeman and J. McKenzie, "SPARK, a confidential web-based template for self and peer assessment of student teamwork: Benefits of evaluating across different subjects," *Brit. J. Educ. Technol.*, vol. 33, no. 5, pp. 551–569, Dec. 2002, doi: 10.1111/1467-8535.00291.

[16] H.-T. Hou, K. Chang, and Y.-T. Sung, "An analysis of peer assessment online discussions within a course that uses Project-based learning," *Interactive Learn. Environments*, vol. 15, no. 3, pp. 237–251, Dec. 2007, doi: 10.1080/10494820701206974.

[17] T. C. Hsu, "Effects of a peer assessment system based on a Grid-based knowledge classification approach on computer skills training," *J. Educ. Technol. Soc.*, vol. 19, no. 4, pp. 100–111, 2016.

[18] G. Hwang and S. Chang, "Facilitating knowledge construction in mobile learning contexts: A bi-directional peer-assessment approach," *Brit. J. Educ. Technol.*, vol. 52, no. 3, pp. 337–357, Jul. 2020, doi: 10.1111/bjet.13001.
[19] A. Jaime, J. M. Blanco, C. Domínguez, A. Sánchez, J. Heras, and I. Usandizaga, "Spiral and project-based learning with peer assessment in a computer science project management course," *J. Sci. Educ. Technol.*, vol. 25, no. 3, pp. 439–449, Jun. 2016, doi: 10.1007/s10956-016-9604-x.
[20] M. Kim and J. Ryu, "The development and implementation of a web-based formative peer assessment system for enhancing students' metacognitive awareness and performance in ill-structured tasks," *Educ. Technol. Res. Develop.*, vol. 61, no. 4, pp. 549–561, Aug. 2013, doi: 10.1007/s11423-012-9266-1.
[21] A. Laflen, "Using eli review as a strategy for feedback in online courses," *Assessing Writing*, vol. 46, no. 1, Sep. 2020, Art. no. 100486, doi: 10.1016/j.asw.2020.100486.
[22] C. Y. Lai, "Training nursing students' communication skills with online video peer assessment," *Comput. Educ.*, vol. 97, pp. 21–30, Mar. 2016, doi: 10.1016/j.compedu.2016.02.017.
[23] X. Liu and L. Li, "Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment," *Assessment Eval. Higher Educ.*, vol. 39, no. 3, pp. 275–292, Apr. 2014, doi: 10.1080/02602938.2013.823540.
[24] S. Loddington, K. Pond, N. Wilkinson, and P. Willmot, "A case study of the development of webpa: An online peer-moderated marking tool," *Brit. J. Educ. Technol.*, vol. 40, no. 2, pp. 329–341, Mar. 2009, doi: 10.1111/j.1467-8535.2008.00922.x.
[25] M. L. Loughry, M. W. Ohland, and D. J. Woehr, "Assessing teamwork skills for assurance of learning using CATME team tools," *J. Marketing Educ.*, vol. 36, no. 1, pp. 5–19, Apr. 2014, doi: 10.1177/0273475313499023.
[26] S. Mader and F. Bry, "Towards an annotation system for collaborative peer review," in *Proc. Int. Conf. Methodol. Intell. Syst. Technol. Enhanced Learn.*, 2020, pp. 1–10, doi: 10.1007/978-3-030-23990-9_1.
[27] E. Panadero and M. Alqassab, "An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading," *Assessment Eval. Higher Educ.*, vol. 44, no. 2, pp. 1–26, Apr. 2019, doi: 10.1080/02602938.2019.1600186.
[28] T. Papinczak, L. Young, M. Groves, and M. Haynes, "An analysis of peer, self, and tutor assessment in Problem-based learning tutorials," *Med. Teacher*, vol. 29, no. 5, pp. e122–e132, Jul. 2007, doi: 10.1080/01421590701294323.
[29] M. J. Prince and R. M. Felder, "Inductive teaching and learning methods: Definitions, comparisons, and research bases," *J. Eng. Educ.*, vol. 95, no. 2, pp. 123–138, Apr. 2006, doi: 10.1002/j.2168-9830.2006.tb00884.x.
[30] F. Sciarrone and M. Temperini, "K-OpenAnswer: A simulation environment to analyze the dynamics of massive open online courses in smart cities," *Soft Comput.*, vol. 24, no. 1, pp. 11121–11134, Aug. 2020, doi: 10.1007/s00500-020-04696-z.
[31] B. Sridharan, M. Muttakin, and D. Mihret, "Students' perceptions of peer assessment effectiveness: An explorative study," *Accounting Educ.*, vol. 27, no. 3, pp. 259–285, May 2018, doi: 10.1080/09639284.2018.1476894.
[32] T. Staubitz, D. Petrick, M. Bauer, J. Renz, and C. Meinel, "Improving the peer assessment experience on MOOC platforms," in *Proc. 3rd ACM Learn. Scale Conf.*, 2016, pp. 389–398, doi: 10.1145/2876034.2876043.
[33] A. Sukstrienwong, "A web-based peer assessment system for assigning student scores in cooperative learning," *TEM J.*, vol. 6, no. 4, pp. 717–725, Nov. 2017, doi: 10.18421/TEM64-10.
[34] D. Sun, N. Harris, G. Walther, and M. Baiocchi, "Peer assessment enhances student learning: The results of a matched randomized crossover experiment in a college statistics class," *PLoS One*, vol. 10, no. 12, Dec. 2015, Art. no. e0143177, doi: 10.1371/journal.pone.0143177.
[35] Y.-T. Sung, K. E. Chang, S. K. Chiou, and H. T. Hou, "The design and application of a Web-based self- and peer-assessment system," *Comput. Educ.*, vol. 45, no. 2, pp. 187–202, Sep. 2005, doi: 10.1016/j.compedu.2004.07.002.
[36] J. J. Suñol, G. Arbat, J. Pujol, L. Feliu, R. M. Fraguell, and A. Planas-Lladó, "Peer and self-assessment applied to oral presentations from a multidisciplinary perspective," *Assessment Eval. Higher Educ.*, vol. 41, no. 4, pp. 622–637, May 2015, doi: 10.1080/02602938.2015.1037720.
[37] T. Tenório, I. I. Bittencourt, S. Isotani, and A. P. Silva, "Does peer assessment in on-line learning environments work? A systematic review of the literature," *Comput. Hum. Behav.*, vol. 64, pp. 94–107, Jul. 2016, doi: 10.1016/j.chb.2016.06.020.
[38] J. W. Thomas, "A review of research on project-based learning," Jan. 2000. [Online]. Available: bie.org/research/study/review_of_project_based_learning_2000
[39] D. Tinapple, L. Olson, and J. Sadauskas, "CritViz: Web-based software supporting peer critique in large creative classrooms," *Bull. Tech. Committee Learn. Technol.*, vol. 15, no. 1, pp. 29–35, Jan. 2013.
[40] A. Vista, E. Care, and P. Griffin, "A new approach towards marking large-scale complex assessments: Developing a distributed marking system that uses an automatically scaffolding and rubric-targeted interface for guided peer-review," *Assessing Writing*, vol. 24, pp. 1–15, Apr. 2015, doi: 10.1016/j.asw.2014.11.001.
[41] Y. Wang, Y. Liang, L. Liu, and Y. Liu, "A multi-peer assessment platform for programming language learning: Considering group non-consensus and personal radicalness," *Interactive Learn. Environments*, vol. 24, no. 8, pp. 2011–2031, 2016, doi: 10.1080/10494820.2015.1073748.
[42] J. Wright, C. Thornton, and K. Leyton-Brown, "Mechanical TA: Partially automated high-stakes peer grading," in *Proc. 46th ACM Tech. Symp. Comput. Sci. Educ.*, 2015, pp. 96–101, doi: 10.1145/2676723.2677278.

Gabriel Badea received the B.Sc. degree in computer science and the M.Sc. degree in information systems for e-business in 2016 and 2018, respectively, from the University of Craiova, Craiova, Romania, where he is currently working toward the Ph.D. degree.

He has been a Teaching Assistant with the University of Craiova since 2018. His major research interests are related to technology enhanced learning and peer assessment in education.

Mr. Badea is the recipient of several scholarships and awards, including a distinguished paper award.

Elvira Popescu (Senior Member, IEEE) received the Ph.D. degree in information and systems technologies from the University of Technology of Compiègne, Compiègne, France and the University of Craiova, Craiova, Romania (double degree), in 2009.

She has been a Faculty Member with the Computers and Information Technology Department, University of Craiova, since 2005, obtaining the position of Full Professor in 2018. She authored and coauthored more than 100 publications, including two books, journal articles, book chapters, and conference papers. In addition, she coedited six journal special issues as well as 18 international conference proceedings. She participated in more than 15 national and international research projects, three of which as Principal Investigator (grant director). Her research interests include technology enhanced learning, adaptation and personalization in web-based systems, learner modeling, computer-supported collaborative learning, learning analytics, and intelligent and distributed computing.

Dr. Popescu is the recipient of several scholarships and awards, including five best paper distinctions. She is actively involved in the research community by serving as Associate Editor for three journals, member of five other journal editorial boards, organizing a series of international workshops in the area of social and personal computing for e-learning (SPeL 2008−2020), and acting as conference chair, program committee chair, and track chair for more than 20 conferences. She currently serves as the Vice Chair for the IEEE Women in Engineering Romania Section Affinity Group and is a board member for the IEEE Technical Committee on Learning Technology (TCLT) and the International Association of Smart Learning Environments (IASLE). She is also a Distinguished Speaker in the IEEE Computer Society Distinguished Visitors Program.