International Journal of Information and Communication Technology Education
Volume 17 • Issue 3 • July-September 2021

Self and Peer E-Assessment:
A Study on Software Usability

Rosalina Babo, ISCAP, Polytechnic of Porto, Portugal
https://orcid.org/0000-0002-8505-8783
Joana Rocha, ISCAP, Polytechnic of Porto, Portugal
Ricardo Fitas, Faculty of Engineering, University of Porto, Portugal
Jarkko Suhonen, University of Eastern Finland, Finland
Markku Tukiainen, University of Eastern Finland, Finland

ABSTRACT

In recent years, universities and other academic institutions have been performing collaborative learning to improve the students' skills, competencies, and learning outcomes. The problem-based learning approach is a learning method where students can work in groups to develop these skills. However, when working in groups, the students are not always being assessed accordingly to their contributions to the work development. Therefore, self and peer assessment are becoming a common practice in academic institutions. With the technology evolution and the emerging of assessment software tools, the evaluators' work is becoming easier. This paper presents seven different tools, as well as their features and functionalities. To evaluate and compare these tools, some parameters are presented and described, based on usability and user experience definitions. It concludes that there is a demand to develop a freeware tool, with parameters not presented in the existing evaluation tools to assist the lecturer in the assessment.

KEYWORDS

Collaborative Work, E-Assessment, Evaluation, Higher Education, Peer Assessment, Self and Peer Assessment, Self-Assessment, Software Tools, Usability, Workgroup

INTRODUCTION

Situations in which students are required to collaborate in small groups to improve their learning is called collaborative learning. This instructional approach involves group exchanges to usually solve or create a project (Johnson, Johnson, Smith, & Smith, 2013) (Hernández-Sellés, Pablo-César Muñoz-Carril, & González-Sanmamed, 2019).

One method of collaborative learning is Problem-Based Learning (PBL). The aim of PBL is to improve the creative thinking skills, problem-solving skills, and learning outcomes of students (Khoiriyah & Husamah, 2018). With this method of learning, the students can work in groups, in order for them to discuss the best way to solve the problem.

When working in groups, students are actively engaged in "group learning to gain content knowledge and problem solving skills", and thus they are "assessed based on their contributions to the group learning" (Masek & Salleh, 2015). Traditionally, the collaborative work is assessed

DOI: 10.4018/IJICTE.20210701.oa5

accordingly to the final project, consequently the members of a group are given the same mark for their collaborative work (Macdonald, 2003). However, Kolmos & Holgaard (2007) affirm that a "variation in assessment practices" have been increasing, since workgroups require group resources, that is knowledge and practice of every member, therefore the assessment should "capture this ability" (Kolmos & Holgaard, 2007). Hence the student-based assessment (self and peer assessment) will "enhance the authenticity and inclusiveness in assessments of PBL" (Masek & Salleh, 2015).

In recent years, there has been an increasing focus on assessment (Tomas, Borg, & McNeil, 2015). Making self and peer assessment is becoming a huge habit among universities and other academic institutions (Tan & Keat, 2005). As this concept has been developed in recent years, also the regular assessment became a part of companies to make a constant analysis of their developed work.

However, according to Clark, Davies & Skeers (2005), the task of "performing a fair and accurate assessment of individual student contributions to the work produced by a team as well as assessing teamwork itself presents numerous challenges". Luaces et al. (2018), affirms that the assessment is subjective and thus, "the help of an intelligent system capable of performing this task is needed". Therefore, with the recent development of technology, there is the possibility of automation, which can make the assessment more efficient and faster. E-assessment is defined as "the use of information and communication technology to mediate any part of the assessment process" (JISC, 2007, cited by (Tomas, Borg, & McNeil, 2015).

The unfairness of individual's assessment was clearly noted by the lectures of Porto Accounting and Business School (ISCAP) who had increased their use of PBL in their classrooms. Since there was no manner to distinguish the workgroup member, the assessment could be biased. Usually, the lecturers would attribute the same mark to the whole group. This practice was unfair, since it most of the times did not corresponded to the actual contribution and performance of each individual.

With individuals' constant motivation for implementing new tools or improving existing ones, to support unbiased assessments, concepts like usability, user experience, assessment and the relation between them are being studied, and the results are being considered and implemented. The application of this concepts results in some online tools which aim to facilitate the evaluators' work.

This article presents examples of online self and peer assessment tools and usability comparative parameters. It is part of a bigger study that aims to understand how to perform fair individual assessments in workgroups, as well as if there is a good methodology to assist the evaluator in their tasks. For that, Design Science Research (DSR) was used, and this study comprehends the rigour cycle. It attempts to understand if the existing assessment frameworks are ideal for distinguishing members of workgroups and if the design of a new assessment framework would be advantageous. It is part of the knowledge base foundations, by comparing existing assessment tools to ensure actual research contributions.

In the next sections, present PBL's characteristics and importance, some definitions of usability and user experience, and explained how these concepts can be fundamental to improve these tools, and thus achieve better results. Then the used methodology will be explained. After that, there will be the discussion on characteristics deemed crucial to the evaluation of tools and the verification of the existence of these in online tools. Finally, the conclusion section will present the features necessary to improve these tools and the motivation for the design of a new assessment tool.

BACKGROUND OF CONCEPTS

The following section presents some relevant concepts when discussing software tools for assessment. These concepts are Problem Based Learning, software tools, and assessment tools parameters.

Problem Based Learning: The Importance of Self and Peer Assessment

The assessment concept has been growing in the recent years. Nowadays, it is frequent to perform self and peer assessment. The self-assessment consists in evaluate the individual work, by having

the person evaluate himself, while the peer assessment measures the collaboration between peers in projects and presentations (Ward and Masgore, 2004 cited by (Daba, Ejersa, & Aliyi, 2017)). These activities can be applied to Project Based Learning (PBL).

The goal of this method is to assist the students to "develop long-term learning skills" (Frank & Barzilai, 2004), such as content, process and problem-solving skills, which can enable them to face the 21st century challenges (Khoiriyah & Husamah, 2018).

According to Frank & Barzilai (2004), with the PBL approach the students are required to carry out a project, proposed by the lecturer, where they have to design a "system based on scientific, technological, social and environmental principles". When doing so, they will explore "important and meaningful questions through a process of investigation and collaboration" (Frank & Barzilai, 2004). In PBL, the students have to perform collaborative work with others, which allows them to reflect in the subjects learned (Kızkapan & Bektaş, 2017).

Considering that the students must work in groups, the importance of self and peer assessment is clearer, since not every member has the same performance (Hall & Buzwell, 2013). While working in groups the students have the "opportunity to observe their peers throughout the learning process and often have a more detailed knowledge of the work of others than do their teachers" (Dochy, Segers, & Sluijsmans, 1999). They have a unique point of view that the lecturer does not have and can also compare their performance to the other members of the group. According to Dochy et al. (1999) involving the students in the assessment process can be "perceived as being valid, reliable, fair and as contributing to a growth in competence.".

Software Tools: Usability and User Experience

Associated with technological advances, software tools to more efficiently perform self and peer assessment of students started to emerge. Related to these tools are the concepts of usability and user experience.

The concept of usability is crucial for the analysis of the interaction of software tools with the users. Most authors (McCall, 1977; Eason, 1984; Makoid, 1985; Shackel and Richardson, 1991; Bevan et al., 1991; Dubey et al, 2010, pp.4723-4729; et al., cited by Đorđević, 2017) introduce this concept to explain user experience relationships and users' products. According to ISO 9241-11:1998 definition, usability is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Apart from this definition of usability also Shackel (2009) defined this keyword as the capability of an object to be used in an easy and effective way by the intended users (Haaksma, de Jong, & Karreman, 2018). Nevertheless, usability can be applied in various subjects, like new mobile applications, public transports, domestic machines or, as this paper introduced, software tools for assessment.

When comparing user experience (UX) with user interface (UI), it can be stated that UX is used in a broader way. This is because when evaluating a software tool, the whole experience with it must be considered and not just the visuals on the screen. For example, when going to a fashion store and seeing clothes that look good, but then when buying some clothing items, the store service is poor. One can say that the overall experience with the store is not good. Certainly, that the look of the UI is important, but the user would be more productive when having all the necessary features, and thus an overall good experience.

In order to understand how the users perceive the tools, usability testing is performed. From these tests, the usability of a tool is "determined by a combination of its features, functionality, visual appeal, and usefulness" (Corrao, Robinson, Swiernik, & Naeim, 2010). According to Costabile, et al. (2005), the tool has great usability when it allows the users to manipulate it efficiently and when it performs the tasks appropriately.

The usability testing also assesses the way that the users experience the tool and determine who are the target audience of it, since the tool must be suitable for the target users. Some tools are not oriented to older audiences, because the senior people have more difficulties in understanding how

the software tools work and its usability is hard to grasp. Luxton-Reilly (2009, p. 21) conducted a systematic review on peer assessment tools, summarizing the available tools and classified them. One major conclusion of this study was that there was a "clear need for more usability studies" on the different tools that exist in the market. Thus, an assessment tool must be capable to assist a range of specific teachers in order to efficiently perform the task of evaluating and marking students (Enriquez, Brito, & Orellana, 2017; Luxton-Reilly, 2009).

Assessment Tools Parameters

There are many software tools with different capabilities that can be used in e-assessment. These tools allow to provide different assessment activities, as well as the "recording of responses, timely feedback, automatic grading, and weighted-average grade calculation" (Abelló Gamazo, et al., 1992).

Literature review shows that the most used key characteristics of the definitions of usability are efficiency, learnability, and satisfaction. According to Đorđević (2017), almost 12.4% of authors refer "the ease of learning" as the crucial attribute for usability, and satisfaction, flexibility, efficiency, and effectiveness, as complements of the usability definition.

A comparative study will be presented in order to analyse the main characteristics of some existent tools. To do this, besides the characteristics previously referred (ease of learning, efficiency, satisfaction and flexibility), the five important parameters of usability, defined by Nielsen, learnability, efficiency, memorability, errors, and users' satisfaction, have to also be taken into account (Nielsen, 1993) (Muqtadiroh, Astuti, Darmaningrat, & Aprilian, 2017).

In order for a tool to have a good usability, it has to be designed according to the characteristics stated above. When a tool follows these, its use is more likely. Therefore, tools used for self and peer assessment must also regard these characteristics, thus both the users, evaluators and evaluated, can have the best experience.

Considering that the students have to self and peer evaluate their work systematically, the most efficiently they can do this, the fastest it is done. In order to achieve that, one can start by shorten the size of surveys, therefore students can answer them quickly. A study done by Rolstad, Adler, & Rydén (2011) states that it is more beneficial to shorten the content of a survey, instead of its length, even though lengthy surveys are more likely to be abridged. Since the response burden is more marked when the survey is full of complex questions than when it is longer, but with "more straightforward response alternatives".

Even though lengthy surveys can deliver more information, they are more time-consuming, which increases the responder's burden and thus their "willingness to participate and/or complete questionnaires" (Dillman et al., 1993, Rogelberg and Luong, 1998, Stanton et al., 2002 cited by (Smits & Vorst, 2007).

According to Van Selm & Jankowski (2006), the rule when designing surveys is "the longer the questionnaire, the less likely people will respond". Which is coincident with Edwards, Roberts, Sandercock, & Frost (2004) study, who affirms that the surveys should be "made as short as possible without compromising the data collection", since the responses may increase when using shorter surveys.

Therefore, in order to respect these statements, the surveys should be short and composed of closed structure questions (McColl et al., 2001, cited by (Mackison, Wrieden, & Anderson, 2010), since it can reduce the respondents' burden by decreasing the content, and at the same time shorten the length of the surveys.

Another important factor to consider in the evaluation of assessment tools is the ease to access these tools by people interested in using them. The development of an assessment tool involves monetary and time costs, therefore the designers and developers will want to monetise their work (Maican & Lixandroiu, 2016). The users will perceive that by paying for the tool, it will have more benefits and will be more powerful for managing the evaluation grades of students (Lubas, 2016).

Nevertheless, there are also other tools which may be open to the community and have the same options and features.

Besides monetary costs, it is also important to highlight time costs. The evaluation tool must be time saving in order to have advantages of usage. Tools with too many functionalities can be time-consuming and some of those functionalities are often not used. The ideal is to have the necessary features for a good user experience.

When the users are free to adjust the parameters of the software tools, this turns the tool more versatile and adaptive. An example of a feature that the evaluator could benefit from, is a feature in which the evaluator can adjust the weight of the assessment given by the students. That means that one evaluator can give, for example, more importance to peer evaluation than to self-evaluation, but another evaluator can think the opposite. That is why the option to adjust the weight of assessments parameters is an added value (Wahid et al., 2017).

Among all the tools, there may exist some that have the most usage, which means they have higher user engagement. The user engagement comprises the initial reaction of the users about the tools and their continuous use over time (Kokil, 2018). Also, the concept of usability is often presented by its efficiency characteristic. However, other characteristics can be featured in the implementation of a well-organised software. Such as emphasizing the most important aspects to have a better guidance (learnability), having a non-monotonous design, then users do not forget certain actions (memorability), a well-implemented system that can recognise the most common errors and solves them to prevent the users from giving up the tool (errors detected), and the best adaptation for users, in order for them to benefit from it (user satisfaction).

METHODOLOGY

This study uses the Design Science Research (DSR) methodology, which focus on the development and improvement of solutions to relevant problems. This methodology has three iterative research cycles: relevance, rigour, and design cycle (Figure 1). The relevance cycle offers the context of the problem. The rigour cycle has the foundations necessary to develop a framework. Then, the design cycle involves both relevance and rigour cycles to design and build a new framework. This cycle is especially relevant since it ensures the development evaluation to identify possible improvements, leading to the development and implementation of alternatives. Since it is iterative, the design and development of a framework will have constant evaluations and redesigns until an ideal solution is reached (Hevner, 2007; Hevner & March, 2003; March & Storey, 2008).

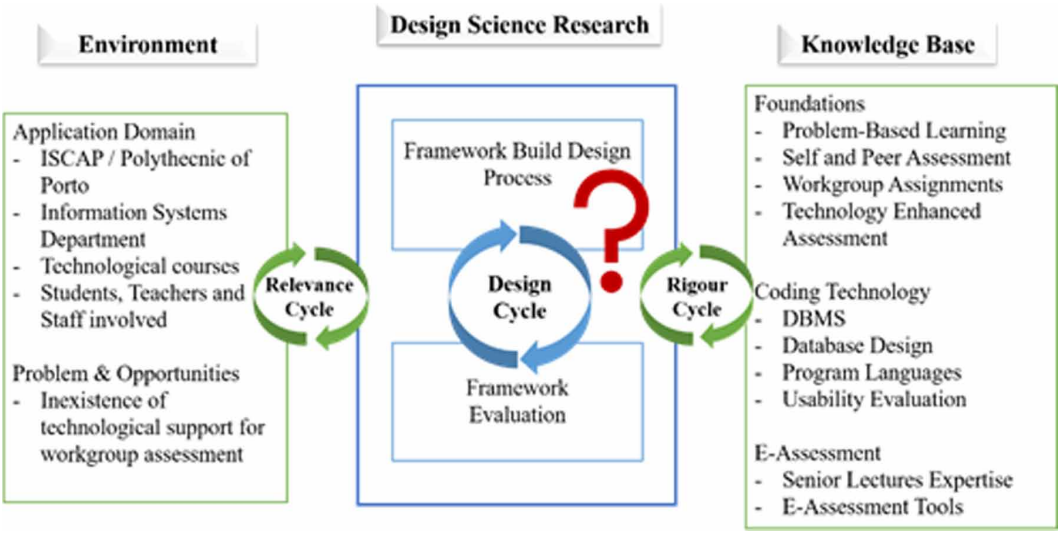
The motivation to study this topic led from the inexistence of technological support for the assessment of workgroups at ISCAP, which stood both as a problem and an opportunity. In order to understand whether there are available solutions to this problem, the present study looks into some of the existing assessment software tools and compares their features and capabilities. Therefore, it presents the knowledge base foundations, which are a part of the rigour cycle. The research questions (RQ) are:

RQ1: *Are the existing assessment frameworks ideal for distinguishing workgroup members?*
RQ2: *Would the design of a new assessment framework be advantageous/necessary?*

There are several tools available in the market that can help the evaluator in their task of assessing students. Therefore, in order to find software tools to compare that could answer the research questions, some criteria had to be delineated for their search and selection:

- To be a software tool to assess workgroups
- Have the capability to perform self and/or peer assessments

Figure 1. Design science research cycles (Adapted from Hevner, 2007)



- To be available online: have a website or any type of web presence
- Have scientific publications referring to them.

The search allowed the finding of several software tools that, although similar, had some differences. These differences may ensure that the different users use the tool that best suits them, and they also encourage the implementation of some ideas for the improvement of the remaining tools.

The assessment software tools selected were: WebPA, TeamMates, InteDashboard, Workshop module, PeerMark, iPeer, and Peergrade. The following subsections will present these tools, their main features, and usability.

WebPA

WebPA is a private software tool that allows teachers to manage groups and assignment marks. Unlike other tools, Loddington, Pond, Wilkinson, & Willmot (2009) published an article about this tool that explains in a simple manner how the students can evaluate their work, by answering a questionnaire that allows easiness and quickness.

After having logged in the system, the evaluators can set some initial parameters like groups and scales' size, can build forms, and can determine the weight of the questions (Figure 1). The specifications for the forms' time are also user dependent and teachers can view the status for each form. That is, they can view if the form is open or closed for students' assessment.

In order for the students to take an assessment, they will have to choose the one they want to answer at the moment, among the current assessments available. After the choice, a new page will be displayed. All of the questions are scaled, by default, from 1 to 5, and the users are able to manage the scale. There are not short or extended questions, and the forms are very quick to answer. This is one of the advantages of WebPA in relation to other tools, since the students must answer the forms without the subjectivism that open questions have.

However, evaluating students based on their answers may not be easy when all of them answer the same in all the questions. This "non-restricted" characteristic can make it impossible to distinguish group members, which is a particularity that can turn the WebPA inefficient.

Another feature that the system does not include is a notification feature or any type of mailing notification. Furthermore, the WebPA system does not display the results directly and although the information given by the system is simple, it becomes repetitive.

Despite this, this tool has excellent features that makes it easy and fast for students and teachers to use and learn (WebPA, n.d.).

Peergrade

Peergrade is a proprietary software tool that can be used in the assessment process, particularly in the review of assignments. Which means that it has a different utility than other assessment software tools, since its main purpose is to assist the lecturer in a continuous and rigorous evaluation. It makes the best connection between students and teachers, since the teacher has a complete overview about students' participation.

This tool has a feature that allows the upload and download of files. Also, the vast amount of automatic statistics algorithms incorporated in this tool demonstrates a rigorous particularity although its entire use is unlikely. The tasks performed in the tool are not quick, because they can sometimes take exhaustive answers. This is an incomparable tool, in relation to the other tools presented here, since its main goal is to submit and review assignments, in opposite to the remaining tools, which aim to provide final grades for students. This tool was also made for a reduced target audience, considering that not every evaluator needs a tool so powerful to evaluate their students.

Indeed, not every assessment tool has the same aims and audiences, and the features of Peergrade allow to ascertain this. There are tools more effective than others, and Peergrade can be very effective if the people who use it are looking for something specific. It is important to reinforce that Peergrade has more options that can interest different individuals (Peergrade - engaging student peer review, n.d.)

iPeer

The iPeer tool is a peer review application where the evaluator can customise and receive students' evaluations. This tool offers three types of peer evaluation, among which the evaluator can choose. Those can be simple, rubrics or mixed. The first consists in the distribution of points among the members of a group. In the second, the students answer a multiple-choice questionnaire. And the last is a combination of the simple and rubrics evaluations.

By having the option to choose among the types of assessment, the evaluator can adjust the assessment according to the carried course, project or group work.

This tool has private access and restrictions for better evaluation are not considered, because there is the possibility for open answers. This can turn the tool subjective and not very capable for adjusting marks (iPeer | Teaching with Technology, n.d.)

TeamMates

TeamMates has particularities that differs it from the other tools. This tool does not have a self-evaluation option and groups can only be formed by filling a form from an MS Excel spreadsheet. Despite this tool having no algorithmic restrictions and that it can support open response questions, where there is great emphasis on the interactions with comments, there is also an option that allows the students to answer objectively, with scaling type questions.

However, the scaling type questions have a different way of working. There is a range of responses in which students evaluate their peers through the increase or decrease of income in relation to the average work of team members. Moreover, weighting is not possible, and the display of the results are not directly shown. There is also the option to send notifications automatically to inform about the open or finished assessment forms, in order to keep students alert for the assessment status.

Nevertheless, TeamMates has a great usability, but the fact of being a private tool and less capable for interpreting the results can turn this into an unusable tool, which makes the users search

for alternative resources ("TEAMMATES - Online Peer Feedback/Evaluation System for Student Team Projects", n.d.).

Peermark

The Peermark is a feedback tool, part of the Turnitin Feedback Studio. Turnitin is an originality checking and plagiarism prevention service that verifies citation mistakes or wrongfully copying.

The goal of Peermark is the revision of assignments done by students. In this tool, the students review the papers of their colleagues by answering free response and scale questions. The review can be done anonymously, and there is an option to add comments and other notes.

Peermark is only comparable with Peergrade, by being similar to it in some aspects. However, this tool does not require any type of paid subscription. Also, Peermark's target audience are the people that only use evaluation occasionally in a course, since its use is very restrictive to a type of evaluation.

Moreover, some information about mail notifications or user restrictions are not clarified, and it does not contain the self-evaluation feature. But like most tools, the report option and open/close form status are part of Peermark ("PeerMarkTM - Guides.turnitin.com", n.d.).

InteDashboard

InteDashboard is a proprietary online software developed for team-based learning. This tool has various features, and although it does not have a self-evaluation feature, it does include a peer evaluation one. This description will be focused on that feature.

The content of the peer evaluation feature is mainly a QA assessment. Hence, instead of the work developed, a group of students test their knowledge about a subject, in order to be able to evaluate each other about their knowledge acquired in that course. For this reason, the questions are not aimed to fairly evaluate. But it is also possible to exchange points of view with the remaining group members in confidentiality. Also, the questionnaires have a timer that can be regulated by the users, in general. Furthermore, to define groups, there is an MS Excel adaptation.

Once again, this tool is not comparable with most of the tools presented here, since it was designed for every academic context in general. InteDashboard has a good usability and shows the ease, the great utility and efficiency of a QA assessment for improving student's capabilities, skills and team work. ("InteDashboardTM - Empowering TBL with Technology | Peer Evaluation", n.d.).

Workshop Module

The Workshop module software is an activity of the Course Development and Management Features from Moodle. Moodle is a free online learning management system which enables educators to create a private website for courses and contributes to the learning.

Workshop is a powerful peer assessment activity that differs from other tools, due to some original aspects, for example tutorials.

This tool is an advanced peer review activity, that can calculate the grades of the students based on their self and peer assessment. Despite of having a non-restricted algorithm incorporated, this software can be very flexible. The users can choose their approach of evaluation, as well as choose the fairness of the assessments' comparison according to their preferences. The evaluation can also be complemented with comments. In addition, it has the possibility to provide open response questions on forms, and those forms can be done for every specific skill.

In terms of usability, the information given by the operational system makes it simple and easy to learn. Workshop module has some examples on how to use it, how to respond to forms, and the evaluation phases are well described. Furthermore, the information about the status of each form is complemented by the status of the different phases, when a form is not answered in its totality.

In comparison to other tools, this one is not complicated to use. It only has more information to guarantee that the user is using the software tool correctly, which improves user experience ("Workshop activity - MoodleDocs", n.d.).

This section presented some software tools, by analysing their features and explaining how they function. It was made some appraisals about the usability of the tools, and how they can be compared with each other. The next section will present a further comparison between them.

Software Tools Comparison

In order to better understand the differences between the software tools previously presented, the Table 1 provides an overview of all the parameters analysed, making it easier to compare the tools. These parameters are drawn from the overall features of those tools.

Parameters Description

In the Assessment Type feature, the table displays what type of assessments are available in the tool, which can be self and/or peer assessment. Evaluating is described in the literature as leaning about learning outcomes (Farrell & Rushby, 2016). It is stated as advantageous for students to develop evaluation skills on themselves (McNamara & O'Hara, 2005, 2008). The ***self-assessment*** is used by the student to describe his individual performance during the development of a project. The importance of self-assessment implementation in the assessment tool is favourable to practice these skills of evaluation on it and that students gain critical thinking about the evaluation. The analysis of the literature says that, in a context of Problem Based Learning, if the members of a workgroup agree to its application, self-assessment have positive effects in the evaluation process (Vanhoof et al., 2009) and, consequently, in the assessment tool.

The ***peer assessment*** is employed in workgroups, thus the students can assign a score to their peers. The feedback give from the other elements have significant influence for the evaluator, to engage him in the group learning (Alias et al., 2015) and for the evaluated, to be more satisficed about their work and his confidence that feedback causes (O'Donovan, 2017).

The ***Survey Adaptability*** is meant to understand if the tool has the option to allow the creation/designing of surveys. This parameter is used to explain the adaptability of the surveys in an assessment tool and it can indicate if there is the possibility to create the questions presented in it. When developing a survey, the authors suggest the user being allowed to edit and create his own survey for the students, to turn the software more flexible and personalized for the teacher. Some literature (Bevan et al., 2016; Rahayu et al., 2018) discuss the importance of "suitability of individualization", one of the seven Indicators of dialogue principles defined in ISO 9241-110, indicating the design of the system is though according to the need of the users. They defend their application, for example, in Open Systems journals (Rahayu et al., 2018).

There are some software tools that also allow the students to configure the ***type of questions*** displayed in the surveys. The questions can be of various types and can be used to better understand the performance of the students. Additionally, the table refers which Question Type the tool has, which can be open, multiple choice or scaling questions. The questions can also be open-response questions or closed-response questions. The questions types can also be according to Syahid (2018), where the different Moodle types of questions to assessment are compared. The study concludes that closed-response questions, such as Random Short-Answer Matching, Multiple Choice Questions and True/False are the preferred of the participants, in terms of usability.

The table also expresses if the tool can ***Generate/Assign Groups***, that is to say if the tool has the possibility to assign members to groups, or if there is another method to achieve that purpose. This feature can be useful when the evaluator intends to perform peer assessment in collaborative work. The availability of generating groups in the software tool is motivated by the importance, not only about tool organization, effectiveness, and flexibility, but also of the assignment and group work assessment. In fact, this importance is well described by Forsell et al. (2020) as improving group and quality of work, by achieving accountability (Johnson & Johnson, 2004) and promoting group processes. It can be important to ease of organization in the perspective of the teacher and for anonymity in students' view, because the other groups cannot view their work, in this way.

Another important parameter is the "***Restricted Scales***". This parameter is used to explain when a tool has a function to restrict the students' responses. It is an important feature, because it allows to provide more accurate data and prevent biased assessments. When a student is not able to grant the same score, when using a scale type evaluation, or give the same answer, at open-ended questions, to every member of their group, the probability of assessment errors decreases. Most of the scales implemented are in percentage or from 1 to 5. For a large number of grades, the evaluation score can be obtained using the average or median, but for a small set, that occurs in the majority of workgroups, the average cannot be effective in the calculation of the individual scores. Also, most students lack in evaluation skills, that can jeopardize the individual and collective marks of the group work (Luaces et al., 2017). Also, there are costs associated to planning and resources that distort the equal distribution of the students' performance (Vossen, 2018). For solving this inconveniences, it is expected the implementation of mathematical score models in software tools, regarding, for example, the sum of all scores must be equal to 1, as suggested in (Luaces et al., 2017).

The existence or non-existence of ***Mailing notifications*** is another parameter considered. It informs if the tool has a way to notify the students when they have to accomplish a task or submit their assessment.

Furthermore, the ***Weighting*** parameter is also important in an assessment tool, since it can be used to specify the importance of each criterion in the assessment process. With this feature, the evaluator can apply different weights in each evaluation to adjust the percentage given to it, and thus provide fair assessments. Self and peer assessment can or cannot be the same weight, as the students have the possibility to inflate their own self-assessment scores, which is not useful for a good assessment. Another weighting mean is related with the different moments of evaluation, or even the weight peer assessment has in relation to the teacher mark (Wahid et al., 2017).

The ***Results*** feature is intended to present if the tool provides the user with the final marks of the assessment process. The Results' Display row expresses if the results are available and displayed after the evaluation is complete (at the time) or if these are only displayed sometime later. Anonymity is defended in the literature (Wahid et al., 2017) as it can have different levels of secrecy. In fact, the anonymity can be positive if it is considered the students' names; however, considering the importance of feedback discussed in peer assessment dimension, the different scores and teacher mark can be displayed as a positive factor, because, in this way, the student has the perception of his work visibility from the teacher and from their colleagues' perspectives.

The ***Reports*** row states if the tool has the option to generate reports of the results given by the tool. Reporting can turn easier the work of the teachers or other professionals, since the software tool can do automatic reports, which reduces substantially their time (Hicks et al., 2018). Once this can be an advantage for numerous groups works and future statistical analysis, the authors understand the importance of this dimension.

The feature named ***Pricing*** establishes if the assessment tool is free, private or proprietary. There are freeware tools, which do not involve any monetary costs and the users can use it freely. And there are tools that are proprietary, which implies a cost for the user, but among these tools, some can have a free trial. Besides these two, there are private tools that can only be used by specific users, or in specific institutions. According to the literature (Maican & Lixandroiu, 2016), in an initial phase of a software development, the costs can be sufficient high in order that few institutions can afford. The importance of Open Source Software (OSS) is a free to use software available for anyone. Since everyone can access this software, the tool can be improved more correctly, in order this type of availability ends to be high feasible. Also, the more complex is the tool, the more expensive it will be (Lubas, 2016).

For simplicity, as one characteristic of software usability, students can be more organized if they know if some task is to be done or not. So, the authors consider be important the existence of the information regarding the opening/closing of tasks. Therefore, the software tool can communicate with users, informing them though pop-ups or messages if they have any task to complete.

Table 1. List of the research articles by the respective authors and parameters

Parameters	Search keywords	Relevant literature
Self-assessment	Self-assessment "software tool"	Borg & Edmett, 2018; Harrison & Murray, 2014; Lubas, 2016; McEvoy et al., 2010; Sleem et al., 2010
Peer assessment	"software for peer" review OR assessment	Abdulla, 2008; Setemen et al., 2020; Sirait & Marlina, 2018
Survey Adaptability	(Usability AND flexibility) OR "Suitability for individualization" "software tools"	Bevan et al., 2016; Rahayu et al., 2018
Generate/ Assign Groups	"software tool" assignment groups	Balderas et al., 2018; Restrepo-Calle et al., 2019
Question type	"question type" importance software tool	Iqbal et al., 2016; Lucia et al., 2009
Restrictions of score	--	Luaces et al., 2017; Vossen, 2018
Mailing notifications	"mail notification" software tool	Iqbal et al., 2016; Lucia et al., 2009
Weighting	"weight marks" software	Wahid et al., 2017; Singh et al., 2016; Alturki, 2016
Display Results	--	Wahid et al., 2017
Reports	"automatic report" software tool	Hicks et al., 2018; Hyung et al., 2003; Safdari et al., 2016; Bauriaud et al., 2018; Raza et al., n.d.
Pricing	--	Lubas, 2016; Maican & Lixandroiu, 2016; Papadakis et al., 2017
Information regarding the opening/ closing of tasks	"software tool" simplicity usability	Osagie et al., 2017; Papadakis et al., 2017

Table 1 suggests some research articles that supports and explain the importance of each parameter authors considered for this study.

Table 2 also has a row on Information regarding the opening/closing of tasks which states if the tool has an option to inform when a task is open or closed. As well as other Observations about the tools.

It is important to note that some of the tools cannot be compared, that is why those parameters are marked with a dash. The information that could not be access, was incomplete or unavailable at the moment, were left blank.

Software Tools Analysis

By analysing the table, and observing what was presented in section 5, it is possible to ascertain the parameters that may be considered common to all tools: Peer Assessment, Survey Adaptability, Generate/Assign Groups, and Results' Reports.

The first parameter is related to all tools, since self and/or peer assessment is the main focus of these tools. Thus, it can be confirmed that the platforms either have self-assessment, peer assessment, or the both types of assessment. The second parameter, Survey Adaptability, states that all tools have the ability to build surveys and that they can be configured by the users. At the Generate/Assign Groups row, it is presented that all tools can generate groups or assign members to groups, which is an important feature in peer assessment. When the platform has a different method to group members, it is normally used a MS Excel spreadsheet. At last, it is relevant to note that all assessment tools can generate results' reports, which can be an important support to the evaluators.

Table 2. Comparison between tools, based on authors researches

Tools		WebPA	Peergrade	iPeer	TeamMates	PeerMark	InteDashboard	Workshop module
Assessment Type	Self	YES	YES		NO	NO	NO	YES
	Peer	YES	YES	YES	YES	YES	YES	YES
Survey Adaptability		YES	YES	YES	YES	YES	YES	YES
Generate/Assign Groups		YES	YES	YES	YES (MS Excel adaptation)	NO	YES (MS Excel adaptation)	YES
Question type		1-5 (default) or teachers scale	QA form (subjective)	Simple / Rubrics/ Mixed	"-/+" share scale	Open-ended/ Scale questions/ Review	QA/ Review	Scale / Rubric / Comments / Number of Errors
"Restrictions of score"		NO	NO		NO	NO	-	NO
Mailing Notifications		NO	YES		YES			NO
Weighting		NO		YES		YES		
Results	Display	NO	YES		NO	YES	YES (at the time)	YES
	Reports	YES	YES	YES	YES	YES	YES	YES
Pricing		Private	Proprietary (Free Trial)	Private	Private		Proprietary	Private
Information regarding the opening/closing of tasks		YES	YES (Notifications)		NO	YES	-	YES (by phases)

Consequently, it is important to analyse the other parameters since they have discrepancies between them. Having stated that, none of the tools has implemented the "Restricted Scales" parameter. This parameter restricts the students' answers, and thus they cannot evaluate the remaining team members in an equal manner. It is an important parameter, considering that it can minimize errors on the attribution of grades.

Another parameter that could be significant to the assessment is the Mailing Notification, once it can assist both the evaluator and the students by notifying them when an assessment is occurring or is about to occur and thus prevent missed assessments. Only two tools (Peergrade and TeamMates) have this feature available.

The Weighing parameter is also available in only two software tools (Peergrade and Workshop module). However, once most tools use open answers, this feature cannot be implemented, since it can only be used in scale type questions. This parameter allows the evaluator to adjust the percentage attributed to each criterion of evaluation. An example of its usefulness is at the assessment of projects that are developed throughout a semester. The evaluator can, with this parameter, attribute a smaller percentage to self/peer assessment at the beginning of the semester, normally when the students are still learning the topics. This percentage can be increased throughout the semester, in order to follow the learning progress of the students.

Moreover, although all tools can generate reports, not all can display them at the end of the assessment. InteDashboard is the only one that displays the results at the time that the evaluation is complete, while the other tools (Peergrade, PeerMark, and Workshop Module) only display the results sometime after the completion of the assessment.

The Pricing parameter is also relevant, since the users may choose the assessment tools according to its pricing. Freeware tools can have more adherence, considering that it does not involve costs to the users. Tools that involve monetary costs are proprietary software, and those may have developers behind the tool who are constantly improving its usability and user interface. There are also private

tools that can be used by specific people/institutions. All the tools analysed are either proprietary or private, none is free. However, despite Peergrade being proprietary, it has a free trial.

The parameter with most discrepancies is the Question Type, because each tool has their own settings. There are scales, QA surveys, rubrics, open-ended questions, comments, among others. There are also tools that are design for review assessment, where the students have to review their colleagues work or provide feedback about it (PeerMark, InteDashboard, and Peergrade).

At the end of the table, it is presented more information regarding the tasks in order to understand how the evaluation process works. Either the tool informs on the opening/closing of tasks, by notifications or by phases, or it does not.

Considering the software tools analysed, most tools only perform peer assessment, with just WebPA, Peergrade, and Workshop Module having both types of assessment. Moreover, among the compared tools, Peergrade and Workshop Module are the only ones that can perform weighted evaluations. This means that there is a distinction between peer and self-evaluation, and/or between assessment criterion.

It is also important to mention that some tools do not display the results of the assessment (WebPA and TeamMates), which leads to a lack of feedback. In contrast, Peergrade displays the results, and it also provides more information about it, which makes it more rigorous.

At last, it is important to refer that the characteristics presented in the table are relevant for the comparison of the different assessment tools. Although, there are more parameters that could be featured, this comparison provided useful information about the existing software tools.

FINAL DISCUSSION AND CONCLUSION

In the recent years, there has been an increase of software tools used in the assessment of individuals in groups, which may be a consequence of the increasing use of collaborative work. As a result, the evaluators are looking for innovative ways to perform these assessments and provide fair and quick feedback.

With this study, we have attempted to find and analyse existing software tools that could perform self and/or peer assessment. It aims to answer the lack of literature and knowledge about the existing software. In addition, it provides an expert user tool's comparison considering different parameters identified previously amongst the literature.

This article analysed some software tools used to evaluate students in self and peer assessment processes. Some of the parameters appraised in these tools were Survey Adaptability, Generate/Assign Groups, "Restricted Scales", Mailing Notifications, Weighting, Results, and Pricing.

It shows that is essential to take into consideration the quickness of the tasks, thus it is not time-consuming. Other important parameters to assess students are "Restricted Scales" and Weighting. Regarding the "Restricted Scales", it is an added value to the tools to implement this parameter, since it can minimise the errors in the attribution of scores, preventing biased assessments and providing more accurate data. The Weighting parameter is significant to provide adjusted evaluations according to the learning progress.

The analysis and comparison conducted in this study allowed to answer RQ1 (*Are the existing assessment frameworks ideal for distinguishing workgroup members?*) by concluding that not all the parameters and characteristics essential in an assessment tool are available. Also, according to the usability tests performed by the authors as expert users, it can be communicated that the tools with a better experience were WebPA, TeamMates, InteDashboard, and Workshop module. Followed by the remaining tools PeerMark, iPeer, and Peergrade.

Concerning RQ2 (*Would the design of a new assessment framework be advantageous/necessary?*), this paper concludes that there is the demand for a freeware software tool supporting self and peer assessment, allowing its use to all the academic community.

In the future, it is important that a tool with all these characteristics, as well as a reinforcement of the usability, such as simplicity, efficiency, and user interaction, should emerge. Some parameters related to weighted assessments, scale restrictions for students, and mainly short questionnaires promoting a faster answer, are all important features that must be considered on the future development of a new free evaluation software tool.

REFERENCES

Abdulla, A. (2008). A critical analysis of mini peer assessment tool (mini-PAT). *Journal of the Royal Society of Medicine*, *101*(1), 22–26. doi:10.1258/jrsm.2007.070077 PMID:18263910

Abelló Gamazo, A., Burgués Illa, X., Casany Guerrero, M., Martín Escofet, C., Quer, C., Rodríguez González, M., & Urpí Tubella, A. (Eds.). (1992). *International Journal of Engineering Education*, *32*.

Alturki, R. A. (2016). Measuring and improving student performance in an introductory programming course. *Informatics in Education*, *15*(2), 183–204. doi:10.15388/infedu.2016.10

Balderas, A., Palomo-Duarte, M., Dodero, J. M., Ibarra-Sáiz, M. S., & Rodríguez-Gómez, G. (2018). Scalable authentic assessment of collaborative work assignments in wikis. *International Journal of Educational Technology in Higher Education*, *15*(1), 1–21. doi:10.1186/s41239-018-0122-1

Bauriaud, M., Kanoun, S., Porot, C., Morel, O., & Berry, I. (2018). *Measurement of Red Cell Mass: Development of new software for standardization of results*. https://posterng.netkey.at/eanm/viewing/index.php?module=viewing_poster&task=&pi=3540

Bevan, N., Carter, J., Earthy, J., Geis, T., & Harker, S. (2016). *New ISO Standards for Usability, Usability Reports and Usability Measures.* In M. Kurosu (Ed.), *BT - Human-Computer Interaction. Theory, Design, Development and Practice* (pp. 268–278). Springer International Publishing.

Borg, S., & Edmett, A. (2018). Developing a self-assessment tool for English language teachers. *Language Teaching Research*, *23*(5), 655–679. doi:10.1177/1362168817752543

Bushell, G. (2006). 2). Moderation of peer assessment in group projects. *Assessment & Evaluation in Higher Education*, *31*(1), 91–108. doi:10.1080/02602930500262395

Choi, B., & Pak, A. (2005). A catalog of biases in questionnaires. *Preventing Chronic Disease*, *2*(1), A13. PMID:15670466

Clark, N., Davies, P., & Skeers, R. (2005). Self and Peer Assessment in Software Engineering Projects. *Proceedings of the 7th Australasian conference on Computing education*, 42, 91-100.

Corrao, N., Robinson, A., Swiernik, M., & Naeim, A. (2010). Importance of testing for usability when selecting and implementing an electronic health or medical record system. *Journal of Oncology Practice / American Society of Clinical Oncology*, *6*(3), 120–124. doi:10.1200/JOP.200017 PMID:20808553

Costabile, M., De Marsico, M., Lanzilotti, R., Plantamura, V., & Roselli, T. (2005). On the Usability Evaluation of E-Learning Applications. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (pp. 6b-6b). IEEE. doi:10.1109/HICSS.2005.468

Daba, T., Ejersa, S., & Aliyi, S. (2017). Student Perception on Group Work and Group Assignments in Classroom Teaching: The Case of Bule Hora University Second Year Biology Students, South Ethiopia—An Action Research. *Educational Research Review*, *12*(17), 860–866. doi:10.5897/ERR2016.3006

Dochy, F., Segers, M., & Sluijsmans, D. (1999). 1). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*(3), 331–350. doi:10.1080/03075079912331379935

Đorđević, N. (2017). Usability: Key characteristic of software quality. *Vojnotehnički glasnik, 65*(2), 513-529.

Edwards, P., Roberts, I., Sandercock, P., & Frost, C. (2004). Follow-up by mail in clinical trials: Does questionnaire length matter? *Controlled Clinical Trials*, *25*(1), 31–52. doi:10.1016/j.cct.2003.08.013 PMID:14980747

Enriquez, E., Brito, J., & Orellana, I. (2017). Evaluating the Usability of Online Social Networks Used by Older People. In *2017 International Conference on Information Systems and Computer Science (INCISCOS)* (pp. 316-322). IEEE. doi:10.1109/INCISCOS.2017.40

Frank, M., & Barzilai, A. (2004). Integrating alternative assessment in a project-based learning course for pre-service science and technology teachers. *Assessment & Evaluation in Higher Education*, *29*(1), 41–61. doi:10.1080/0260293042000160401

Haaksma, T., de Jong, M., & Karreman, J. (2018). Users' Personal Conceptions of Usability and User Experience of Electronic and Software Products. *IEEE Transactions on Professional Communication*, *61*(2), 116–132. doi:10.1109/TPC.2018.2795398

Hall, D., & Buzwell, S. (2013). The problem of free-riding in group projects: Looking beyond social loafing as reason for non-contribution. *Active Learning in Higher Education*, *14*(1), 37–49. doi:10.1177/1469787412467123

Harrison, Y. D., & Murray, V. (2014). The Effect of an Online Self-Assessment Tool on Nonprofit Board Performance. *Nonprofit and Voluntary Sector Quarterly*, *44*(6), 1129–1151. doi:10.1177/0899764014557361

Hernández-Sellés, N., Muñoz-Carril, P-C., & González-Sanmamed, M. (2019). Computer-supported collaborative learning: An analysis of the relationship between interaction, emotional support and online collaborative tools. *Computers & Education, 138*, 1-12.

Hicks, S. A., Eskeland, S., Lux, M., De Lange, T., Randel, K. R., Pogorelov, K., Jeppsson, M., Riegler, M., & Halvorsen, P. (2018). Mimir: An automatic reporting and reasoning system for deep learning based analysis in the medical domain. *Proceedings of the 9th ACM Multimedia Systems Conference, MMSys 2018*, 369–374. doi:10.1145/3204949.3208129

Hyung, Y. E., Vissers, D. R., & Amine, K. (2003). Flame-retardant additives for lithium-ion batteries. *Journal of Power Sources*, *119–121*, 383–387. doi:10.1016/S0378-7753(03)00225-8

Intedashboard™ - Empowering TBL with technology | peer evaluation. (n.d.). Retrieved from https://www.intedashboard.com/features/peer-evaluation/

iPeer | Teaching with Technology. (n.d.). Retrieved January 14, 2019, from https://lthub.ubc.ca/guides/ipeer/

Iqbal, M. S., Sadiq, M., Rehman, A., & Khan, T. (2016). Framework for the Development of Automated Inspection Tools. *International Journal of Computer Science and Network Security*, *16*(1), 16–26. http://www.ijcsns.org/

Johnson, D., Johnson, R., Smith, K., & Smith, K. (2013). Cooperative Learning: Improving University Instruction By Basing Practice On Validated Theory. *Journal on Excellence in University Teaching*.

Khoiriyah, A., & Husamah, H. (2018). Problem-based learning: Creative thinking skills, problem-solving skills, and learning outcome of seventh grade students. *Jurnal Pendidikan Biologi Indonesia*, *4*(2), 151–160. doi:10.22219/jpbi.v4i2.5804

Kızkapan, O., & Bektaş, O. (2017). The Effect of Project Based Learning on Seventh Grade Students' Academic Achievement. *International Journal of Instruction*, *10*(1), 37–54. doi:10.12973/iji.2017.1013a

Kokil, U. (2018). *The Impact of Visual Aesthetic Quality on User Engagement during Gameplay*. Academic Press.

Kolmos, A., & Holgaard, J. (2007). *Alignment of PBL and Assessment*. Aalborg Universitet.

Lejk, M., & Wyvill, M. (2001). The Effect of the Inclusion of Selfassessment with Peer Assessment of Contributions to a Group Project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education*, *26*(6), 551–561. doi:10.1080/02602930120093887

Loddington, S., Pond, K., Wilkinson, N., & Willmot, P. (2009). A case study of the development of WebPA: An online peer-moderated marking tool. *British Journal of Educational Technology*, *40*(2), 329–341. doi:10.1111/j.1467-8535.2008.00922.x

Luaces, O., Díez, J., Alonso-Betanzos, A., Troncoso, A., & Bahamonde, A. (2017). Content-based methods in peer assessment of open-response questions to grade students as authors and as graders. *Knowledge-Based Systems*, *117*, 79–87. doi:10.1016/j.knosys.2016.06.024

Luaces, O., Díez, J., & Bahamonde, A. (2018). A peer assessment method to provide feedback, consistent grading and reduce students' burden in massive teaching settings. *Computers & Education*, *126*, 283–295. doi:10.1016/j.compedu.2018.07.016

Lubas, D. G. (2016). System and software cost correlation to reliability. *2016 Annual Reliability and Maintainability Symposium (RAMS)*, 1–6. doi:10.1109/RAMS.2016.7448052

Lucia, A. D., Fasano, F., Scanniello, G., & Tortora, G. (2009). Evaluating distributed inspection through controlled experiments. *IET Software*, *3*(5), 381–394. doi:10.1049/iet-sen.2008.0101

Macdonald, J. (2003). Assessing online collaborative learning: Process and product. *Computers & Education*, *40*(4), 377–391. doi:10.1016/S0360-1315(02)00168-9

Mackison, D., Wrieden, W., & Anderson, A. (2010). Validity and reliability testing of a short questionnaire developed to assess consumers' use, understanding and perception of food labels. *European Journal of Clinical Nutrition*, *64*(2), 210–217. doi:10.1038/ejcn.2009.126 PMID:19904290

Maican, C., & Lixandroiu, R. (2016). A system architecture based on open source enterprise content management systems for supporting educational institutions. *International Journal of Information Management*, *36*(2), 207–214. doi:10.1016/j.ijinfomgt.2015.11.003

Masek, A., & Salleh, H. (2015). Self, Peer and Teacher Assessments in Problem Based Learning: Are They in Agreements? *Procedia: Social and Behavioral Sciences*, *204*, 309–317. doi:10.1016/j.sbspro.2015.08.157

McEvoy, F. J., McEvoy, P. M., & Svalastoga, E. L. (2010). Web-Based Teaching Tool Incorporating Peer Assessment and Self-Assessment: Example of Aligned Teaching. *AJR. American Journal of Roentgenology*, *194*(1), W56–W59. doi:10.2214/AJR.08.1910 PMID:20028891

Muqtadiroh, F., Astuti, H., Darmaningrat, E., & Aprilian, F. (2017). Usability Evaluation to Enhance Software Quality of Cultural Conservation System Based on Nielsen Model (WikiBudaya). *Procedia Computer Science*, *124*, 513–521. doi:10.1016/j.procs.2017.12.184

Osagie, E., Waqar, M., Adebayo, S., Stasiewicz, A., Porwol, L., & Ojo, A. (2017). Usability Evaluation of an Open Data Platform. *Proceedings of the 18th Annual International Conference on Digital Government Research*, 495–504. doi:10.1145/3085228.3085315

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2017). Designing and creating an educational app rubric for preschool teachers. *Education and Information Technologies*, *22*(6), 3147–3165. doi:10.1007/s10639-017-9579-0

Peergrade - engaging student peer review. (n.d.). Retrieved from https://www.peergrade.io/?nab=1&utm_referrer=https%3A%2F%2Fwww.google.pt%2F

PeerMark™ - Guides.turnitin.com. (n.d.). Retrieved January 14, 2019, from https://guides.turnitin.com/01_manuals_and_guides/instructor_guides/turnitin_classic_(deprecated)/23_peermark

Rahayu, I. S., Hartanto, R., Wibawa, A. P., Meirawan, D., & Nugraha, E. (2018). *Student interaction with Open Journal System based on ISO 9241-110*. 10.2991/miseic-18.2018.21

Raza, M., & Faria, J. (n.d.). *Helping software engineering students analyzing their performance data: tool support in an educational environment*. Retrieved August 3, 2020, from https://ieeexplore.ieee.org/abstract/document/7965317/

Restrepo-Calle, F., Ramírez Echeverry, J. J., & González, F. A. (2019). Continuous assessment in a computer programming course supported by a software tool. *Computer Applications in Engineering Education*, *27*(1), 80–89. https://doi.org/10.1002/cae.22058

Rolstad, S., Adler, J., & Rydén, A. (2011). 12 1). Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health*, *14*(8), 1101–1108.

Safdari, R., Mokhtaran, M., & Tahmasebian, S. (2016). Designing and Implementation of Retina Image Drawing System and Automatic Report Generation from Retina Examinations. *Acta Informatica Medica*, *24*(5), 308–312. https://doi.org/10.5455/aim.2016.24.308-312

Seaman, C. (2008). Qualitative Methods. In C. Seaman, F. Shull, J. Singer, & D. Sjøberg (Eds.), *Guide to Advanced Empirical Software Engineering* (pp. 35–62). Springer.

Setemen, K., Dewi, L. J. E., & Purnamawan, I. K. (2020). Open source software implementation in peer assessment. *Journal of Physics: Conference Series*, *1516*(1). https://doi.org/10.1088/1742-6596/1516/1/012042

Shackel, B. (2009). Usability – Context, framework, definition, design and evaluation. *Interacting with Computers*, *21*(5-6), 339–346.

Shull, F., & Feldmann, R. (2008). Building Theories from Multiple Evidence Sources. In F. Shull, R. Feldmann, F. Shull, J. Singer, & D. Sjøberg (Eds.), *Guide to Advanced Empirical Software Engineering* (pp. 337–364). Springer London.

Shull, F., Singer, J., & Sjøberg, D. (2008). *Guide to Advanced Empirical Software Engineering*. Springer London.

Singh, I., Sabitha, A. S., & Bansal, A. (2016). Student performance analysis using clustering algorithm. *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence),* 294–299. 10.1109/confluence.2016.7508131

Sirait, J. B., & Marlina, L. (2018). Using Instagram As a Tool for Online Peer-Review Activity in Writing Descriptive Text for Senior High School Students. *Journal of English Language Teaching*, *7*(1), 291–302. http://ejournal.unp.ac.id/index.php/jelt/article/view/9771/pdf

Sleem, H., Abdelhai, R. A. A., Al-Abdallat, I., Al-Naif, M., Gabr, H. M., Kehil, E.-T., & Sadiq, B. (2010). Development of an Accessible Self-Assessment Tool for Research Ethics Committees in Developing Countries. *Journal of Empirical Research on Human Research Ethics, 5*(3), 85–98. 10.1525/jer.2010.5.3.85

Smith, M., & Karpicke, J. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory (Hove, England)*, *22*(7), 784–802.

Smits, N., & Vorst, H. (2007). Reducing the length of questionnaires through structurally incomplete designs: An illustration. *Learning and Individual Differences*, *17*(1), 25–34.

Tan, K., & Keat, L. (2005). Self and peer assessment as an assessment tool in problem-based learning. *Problem-Based Learning: New Directions and Approaches*, 162-175.

Teammates - Online Peer Feedback/Evaluation System for Student Team Projects. (n.d.). Retrieved from https://teammatesv4.appspot.com/

Tomas, C., Borg, M., & McNeil, J. (2015). E-assessment: Institutional development strategies and the assessment life cycle. *British Journal of Educational Technology*, *46*(3), 588–596.

Van Selm, M., & Jankowski, N. (2006). Conducting Online Surveys. *Quality & Quantity*, *40*(3), 435–456.

Vossen, P. H. (2018). *Distributive Fairness in Educational Assessment.* In V. E. Balas, L. C. Jain, & M. M. Balas (Eds.), *BT - Soft Computing Applications* (pp. 381–394). Springer International Publishing.

Wahid, U., Chatti, M. A., Anwar, U., & Schroeder, U. (2017). A multi-dimensional peer assessment system. *CSEDU 2017 - Proceedings of the 9th International Conference on Computer Supported Education, 1*, 683–694. 10.5220/0006390006830694

Web, P. A. (n.d.). Retrieved January 14, 2019, from http://webpaproject.lboro.ac.uk/

Workshop activity - MoodleDocs. (n.d.). Retrieved January 14, 2019, from https://docs.moodle.org/35/en/workshop_activity

Rosalina Babo is a Coordinator Professor of Information Systems Department, School of Accounting and Administration of Polytechnic of Porto (ISCAP/PPorto), Portugal. Since the year 2000 she is the head of the Information Systems Department and for about 12 years act as a member of the university scientific board. Rosalina's international recognition was improved with the opportunity to be a Visiting Professor at several universities in different countries namely Belgium (KU LEUVEN), Croatia (University of Split), Kosovo (University of Prishtina), Latvia (Latvia University of Agriculture). Rosalina was one of the founders of CEOS.PP (former CEISE/STI) research centre and its director for 5 years. Rosalina has served on committees for international conferences, and acts as a reviewer in scientific journals. As a book editor she collaborates with publishers such as Elsevier, Springer and IGI Global in the fields of Data analyses in Social Networks and eLearning. Having several published papers her main areas of research are e-learning, e-business, Internet applications focusing on usability and social networks.

Joana Rocha concluded her bachelor's degree in Administrative Assistance and Translation at the Porto Accounting and Business School/ Polytechnic of Porto (ISCAP/IPP), in 2018. She is now a graduate student of Digital Communication Assistance. During her bachelor's degree she intern in the research centre CEOS.PP under Professor Rosalina Babo supervision, providing support in research tasks, which led to her interest in research. In addition, Joana also did an internship with the Polissema - A Journal of Letters of ISCAP, assisting in the edition and administrative work of the 2018 issue of the journal. Besides these activities, she volunteered and supported the realisation of several conferences and events.

Ricardo Fitas is a 20 years old student who studies Mechanical Engineering at the Faculty of Engineering of University of Porto. Interested by science and projects development with high social and technological impact, Ricardo is passionate by improving and developing soft and technical skills every day and helping other students to grow up. He was also a high competition athlete. Despite of the academic field, he gained knowledge in robotic, programming, and user experience. By developing algorithms for images segmentation, optimization, mathematical modulation, statistical analysis, and MATLAB code, Ricardo has gained recognition in Mathematical field. During his freshman year as a university student, he started his research interest, having already presented scientific work in conferences and written for journals. As a student member of IEEE Student Branch of the University of Porto, he participated in important conferences as volunteer, and also lectured workshops for academic community.

Jarkko Suhonen (PhD) holds a research manager position at the School of Computing, University of Eastern Finland. He received his MSc (Computer Science) from the University of Joensuu, Finland in 2000 and his PhD (Computer Science) from the University of Joensuu in 2005. Dr. Suhonen has published over 100 peer-reviewed articles in scientific journals, conferences, workshops, and chapters of books. His research interests include online and blended learning, design science research, computing education, and ICT for development. He has served on committees for international conferences and special issues of journals, and acted as a reviewer in several scientific journals.

Markku Tukiainen (PhD) has been a Professor in Computer Science in the Faculty of Science and Forestry, the University of Eastern Finland since 2004. He earned his Ph.D. in Computer Science from the University of Joensuu, in 2001. Professor Tukiainen is a leader of the research group Interactive Technologies at the School of Computing. Tukiainen has conducted empirical studies in human visual attention, the psychology of programming, and games studies, and has extensive experience with human-computer interaction research, software engineering, and digital accessibility. Professor Tukiainen has published over 90 scientific publications and participates in national standardization work in Finland (through Finnish Software Measurement Organization FiSMA ry) and at an international level (trough ISO International Organization for Standardization). He has been involved in the work of ISO JTC1/SC7 Software and Systems Engineering under Working Groups WG6 Software product Quality Requirements and Evaluation (SQuaRE) and WG20 Standardization of the Software Engineering Body of Knowledge. Currently, he is a co-editor of ISO 25024 Software Product Quality in Use Measurement standard.