

PrESTO

A peer review and peer assessment system with incentives for high quality learning

Bots, Pieter; Els Van Daalen, C.; Dopper, Sofia; Westink, Joosje

DOI

[10.1109/ITHET46829.2019.8937379](https://doi.org/10.1109/ITHET46829.2019.8937379)

Publication date

2019

Document Version

Final published version

Published in

2019 18th International Conference on Information Technology Based Higher Education and Training, ITHET 2019

Citation (APA)

Bots, P., Els Van Daalen, C., Dopper, S., & Westink, J. (2019). PrESTO: A peer review and peer assessment system with incentives for high quality learning. In *2019 18th International Conference on Information Technology Based Higher Education and Training, ITHET 2019* Article 8937379 (2019 18th International Conference on Information Technology Based Higher Education and Training, ITHET 2019). IEEE. <https://doi.org/10.1109/ITHET46829.2019.8937379>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

PrESTO: A Peer Review and Peer Assessment System with Incentives for High Quality Learning

Pieter Bots
Faculty of Technology, Policy and
Management
Delft University of Technology
Delft, The Netherlands
P.W.G.Bots@tudelft.nl

C. Els van Daalen
Faculty of Technology, Policy and
Management
Delft University of Technology
Delft, The Netherlands
C.vanDaalen@tudelft.nl

Sofia Dopper
Education and Student Affairs
Delft University of Technology
Delft, The Netherlands
S.M.Dopper@tudelft.nl

Joosje Westink
Faculty of Policy, Technology and
management
Delft University of Technology
Delft, The Netherlands

Abstract— The PrESTO system is a peer review and peer assessment tool that allows large numbers of students to practice open-ended assignments, while keeping the workload of the teaching staff manageable. PrESTO was originally developed for a quantitative modeling course, but is applicable to any course with open-ended assignments that can be divided into successive steps. The software organizes the peer review and peer assessment workflow, and guides the students through the activities required in each step: review, give feedback on and assess the work of an anonymous predecessor, and then improve and extend it with a next step.

Keywords—peer review, peer assessment, open-ended assignments, software tool

I. INTRODUCTION

Courses that aim at attaining an application level of learning typically require considerable practice from students to become competent. For highly structured subjects like, e.g., algebra, such application practice assignments may have a single right answer, but for many subjects students will have to work on open-ended assignments that result in reports or essays. Providing feedback on and assessing such assignments requires interpretation and judgment, which makes courses very teacher-intensive.

This issue is usually solved by asking students to work in project groups, but this introduces other problems, such as free rider behavior [1]. Alternatively, teachers can mobilize their students to provide such feedback in one or more rounds of peer review, and ultimately even grade the work through peer assessment [2, 3]. However, students tend to put only limited effort into their review tasks [3, 4, 5]. Making review quality count towards the students' final grade typically would require the teacher to assess the reviews, which would mean even more work.

Faced with the challenge to provide students with sufficient individual practice and feedback, while at the same time keeping the teaching burden manageable, we have developed the “project relay” way of working and the supporting software system PrESTO.

During the ITHET conference, we will give a hands-on demonstration of the software. Our objective with this paper is to explain how the PrESTO system works, how it can be used by students and teachers, and what our experiences are in education so far. We also indicate how others can use the PrESTO system and where the software can be obtained.

II. THE PRESTO SYSTEM

A. The “project relay” method

The assignments students carry out in the PrESTO system comprise a number of consecutive steps. Students conduct each of these steps on a different case or topic. The work in each step is peer reviewed. However, an important difference between a normal peer review system (cf. [6] for an overview) and the PrESTO system is that the reviewer, after providing feedback to his/her predecessor, assumes “co-authorship” of the work by improving it before extending it by performing his/her “own” step. Thus, students pass on their work to a fellow student like runners in a relay race pass on the baton to their team mate who will run the next leg. However, in analogy to double-blind peer review in science, students do not know the identity of their predecessors and successors.

In each step (except for the first, where 1-3 do not apply) a student has to conduct the following activities:

1. study the work which has been submitted to the PrESTO system by the predecessor,
2. provide constructive feedback, and then rate the work of the predecessor on a 5-point scale,
3. improve the work,
4. extend it by adding their “own” step,
5. upload the extended work to the PrESTO system.

Fig. 1 illustrates this “project relay” by showing how six students work in sequence on a case, passing on their work. In this example, a quantitative systems modeling assignment is divided into six steps that follow on from each other, starting with a research question and ending in conclusions that can be drawn from experimenting with the model. A relay assignment comprising N steps requires at least N different cases/topics to work on to ensure that students immerse themselves in their predecessors' work.

In the first step, each student is randomly assigned one case. When a student submits his/her work on the first step, this work is assigned to another student who just submitted step 1 on another case. This student reviews and improves the work, and adds step 2 to the case, after which it goes to a next student. This next student then reviews and improves the work, adds step 3, and so on. The dotted arrows in Fig. 1 show how case A is passed on from student to student, while the horizontal lines show that students can start at different times and then “run” the relay at different speeds.

Six or more different cases: **A B C D E F G H**

Six steps:

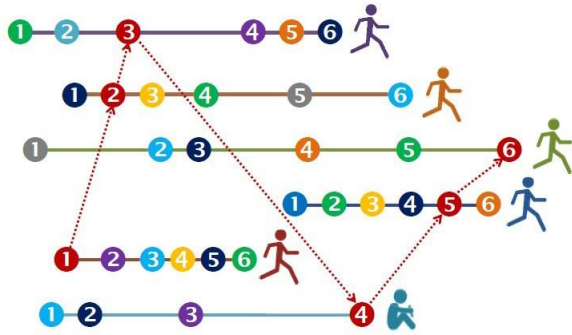


Fig. 1. Diagram showing students going through six different steps

At the end of the process, students will have worked on N different cases (in our example, $N=6$), while practicing N times with step 1, $N-1$ times with step 2, etc. because giving feedback and improving the work always relates to all previous steps. For a video explanation of the method, see <https://youtu.be/SGm-DstdElk>.

The PrESTO system allows students to work at their own pace, as long as they complete all steps before the final deadline. Dropout is not a problem, as students do not obtain work on step x from a fellow student until after they have submitted their own step x . Thus, the “stock” of work is not depleted by dropouts. Work is assigned “just in time” by an algorithm that takes into account (a.o.) the availability of work per step per case and the time since submission (so that fast runners do not have to wait too long on feedback). If for a student wanting to start with step x no unassigned work for step $x-1$ on a case not already worked on by this student is available, PrESTO will make a “clone” of suitable work already assigned to a successor. In this way, students never have to wait unless they really are “frontrunners” in the sense that no other student has submitted work for the previous step. In such rare cases, students are offered the option to either wait, or to continue working on their own work on the previous step. This work will be “cloned”, so that the “original” will still be assigned to another student.

PrESTO features specific incentives to ensure that students put enough effort into the assignments. Students know that their work will be reviewed and assessed by peers in every step, and that this assessment will count towards their final grade. To stimulate critical review as well as improvement of the predecessors’ work, PrESTO uses a differential scoring system (on a 1 to 5 \square scale). This means that students earn points if they receive more stars from their successor than they gave to their predecessor (or if they maintain a score of 4 or 5 \square), and lose points if they receive fewer stars than they gave (or if they do not improve on a score of 1 or 2 \square). This effectively motivates students to do their work well, but it also stimulates students to give their predecessor fewer stars than is merited. To discourage such self-interested harsh scoring, PrESTO offers students who feel that they have been unfairly assessed the option to appeal. When they do so, an instructor will judge the peer assessment and adjust the score if needed. In case of overly

harsh scoring, the successor will incur a penalty point, but students who appeal for no good reason may also lose points.

Even with these incentives, it may happen that students receive work from a predecessor that is below par and would require a disproportionate amount of time to improve. For these situations, instructors can configure PrESTO so that it allows students to reject such work. Students who, based on their review, assess their predecessor’s work with only 1 \square can then – if they so choose – immediately be assigned work of a new predecessor. Since the predecessor whose work is rejected can appeal this assessment, students cannot lightly reject work merely to obtain better work to build on, as they would then risk a penalty from their instructor.

B. The software

The PrESTO software completely automates the peer review workflow, and is available under an MIT open source license. The application runs in all modern browsers on most devices (thanks to the responsive Semantic UI platform). It has been programmed in Python, building on the Django platform. The front-end of the software guides students through the different steps of an assignment. The back-end of the software allows instructors to design different templates that define what a student has to do in every step, develop sets of case descriptions, and manage the whole peer review process (enroll students, assign cases and review tasks, monitor progress, grade students). Students can upload their work as documents of the types specified by the instructor, and a “smart” plagiarism scanner detects where a student’s work contains snippets from submitted work other than that of their “legitimate” predecessors.

C. The student interface

The student interface leads students through the activities that have to be conducted in each of the different steps, i.e., upload own work, start next step, download predecessor’s work, and review and assess predecessor’s work. These activities are always carried out in the same sequence, so that students only see information about the activity they are currently working on, and cannot proceed to the next activity unless they have completed the current one. Meanwhile, students can see (and respond to, and possibly appeal) the review and assessment of their own work as soon as their successor submits his/her work. Predecessors can also download the work of their successor to see whether/how their successor has improved their work. At all times, students can look in their personal history to see their own work, received reviews, appeal decisions, etc.

Fig. 2 shows sections of the screens that students see when they are engaged in an activity. At all times, students can see section \square showing the upcoming deadline, a progress bar that shows how far students are in the process, and a chart depicting the progress of all participants over time. In the progress bar, the steps are numbered and represented by the larger circles. The activities within the steps are indicated by the smaller circles. As the student progresses, the part of the bar to which the student has progressed is shown in grey and the relevant circles are filled in. The same type of activity always has the same color, e.g., \square plum = upload work, \square blue = download predecessor’s work, \square orange = review and assess, and \square green = proceed to the next step. This holds for the progress bar as well as for the other parts of the interface (buttons, menus, and headers).

At the start of each step, students receive a case description, followed by instructions for the task at hand. To improve the work and to add their own step, students continue to work in the file(s) uploaded by their predecessor. The files are anonymized (by removing metadata) when they are downloaded. Students enter reviews into the form (designed by the instructor, q.v.) using a rich text editor.

At the end of the review activity, students are asked to assess the predecessor's work on a 1-5 star scale. Predecessors can indicate what they think of the given review and rating on a three point scale (☐ - ☐ - ☐). If they want to file an appeal, they can click the appeal button and write a rebuttal, which then will be reviewed by the instructor.

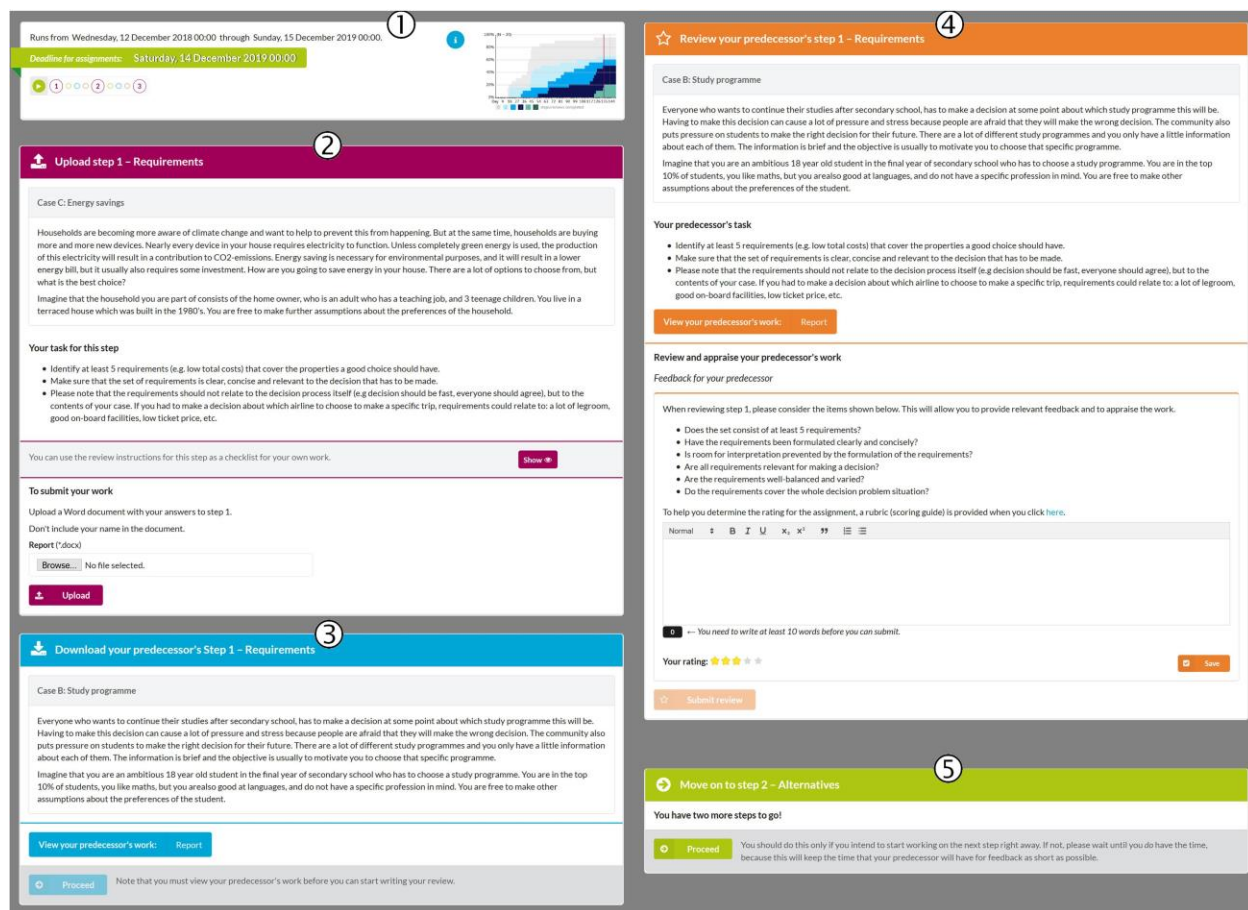


Fig. 2. Sections of PrESTO screens that students interact with during a project relay

D. The instructor interface

The instructor interface allows instructors to define and manage the complete peer review process. Instructors need to complete a template in which they define the number of steps and describe the activities that must be carried out in each step. Fig. 3 shows an example of a screen in which a three step assignment has been defined. Instructors must also define case sets. A set can contain any number of cases, albeit preferably larger than the number of steps. Cases must have a title and description, and may include attachments that students can download.

Instructors also need to specify date and time for the start and final deadline, the number of final reviews, and whether students will receive badges after completing a step and/or a letter of acknowledgement for completing all steps. When they start, students must commit to the “rules of the game”. The instructor can monitor their progress. Fig. 4 shows part of the monitoring screen of a 6 step relay with 2 final reviews. The first column shows students, their progress, when they were last active, and the plagiarism scan (small circles). By clicking on the name of the student, the

instructor can view the student's “relay history”, i.e., all the work the student has submitted, the reviews the student has written and received, the response of other students, etc.

The second column shows the score (star balance plus penalty points) of the students. After a relay, these scores are converted to grades using a scale defined by the instructor after benchmarking the work with the lowest and highest scores. The last two columns show the reviews the students have received and the reviews they have given. The colors of the circles indicate the star rating of the step (red = 1 ☐, dark green = 5 ☐) while the letters in the circles indicate the case the student worked on. The icon (☐ - ☐ - ☐) below a circle shows whether the reviewed student was happy with the review and rating. A hand below a circle indicates that the student filed an appeal. When adjudicated, its color reflects the instructor's star rating, while its pointing direction indicates who incurred a penalty point (right: too harsh assessment by the successor; left: unwarranted appeal by the predecessor). When moving the mouse over these circles and icons, the instructor can see the name of the student involved and the date/time of the event.

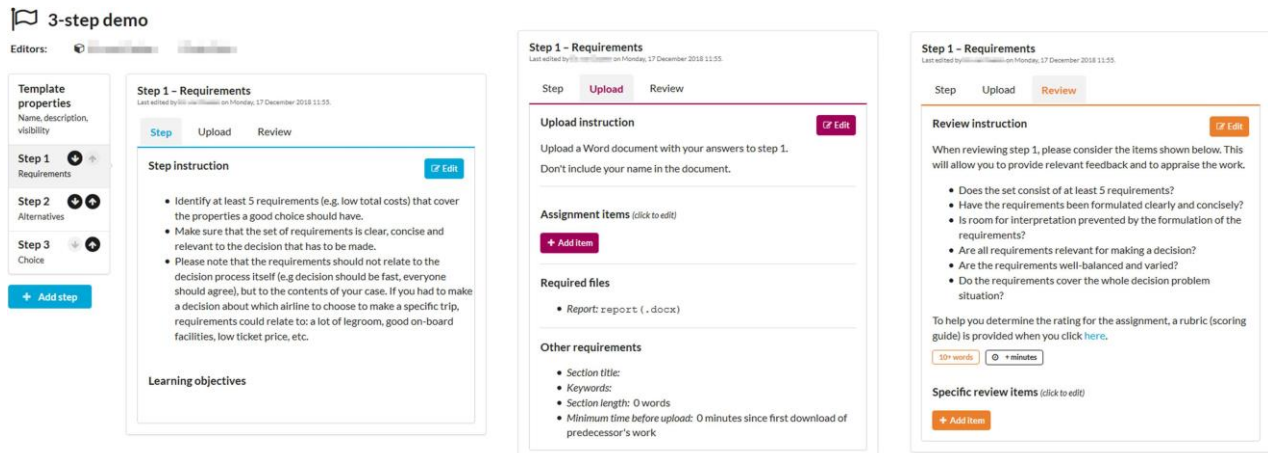


Fig. 3. PrESTO screens for defining the steps of a relay, including instructions for uploading and defining a rubric for review

Name	Score	Received reviews	Given reviews
36% Tuesday, 11 December 2018 12:40	2.0	C A	A E
100% Wednesday, 19 December 2018 17:29	3.0	C D D F E B H H	D F E B H G A
95% Wednesday, 19 December 2018 16:53	4.0	E A F D C H H	A F D C H G
95% Monday, 17 December 2018 17:21	5.0	E C F D G A A	C F D G A H
36% Wednesday, 12 December 2018 11:49	-0.5	E D	D G
73% Monday, 17 December 2018 01:04	4.0	D A E E H H F	A E E H

Fig. 4. PrESTO screen to monitor an ongoing project relay

III. USE IN EDUCATION

The PrESTO system has been used in different ways in various courses taught at Delft University of Technology. Relays are mainly used to practice application of knowledge that is taught in a course, and hence often run in parallel with lectures on theory. We originally developed the relay way of working for a first year bachelor course on quantitative modeling (with a bachelor cohort of between 200 and 300 Dutch students). We have been using and improving PrESTO in this course (with three relays) every year since 2013. From 2017 onwards, we have also been using it in a first year master course on policy analysis with approximately 40 international students. In order to test whether this way of working is generally applicable, we also tested a non-automated version of the relay way of working in a high school, where students had to write an essay. Students wrote a part of the essay and passed this on to each other and wrote the next part, and in that way built up the essay.

Evaluation of PrESTO

The courses in which the PrESTO system has been used have been evaluated using the regular course surveys in

2017-2018, supplemented with specific questions about the peer review process and using data which is captured by the PrESTO system itself, such as timing and performance. Students are of the opinion that they learn a lot from this way of working and that it is a good way to practice. At the same time, they also find the workload of the relay to be high. The experienced work load is on the one hand recognizable, since they first have to improve the work of their predecessor before they can work on a new step (primary task). On the other hand, students also tend to procrastinate, waiting to do the work until the deadline of the final step approaches instead of spreading the work evenly over two weeks.

Students find the process of working in relay fashion to be quite stressful. Although some stress can be functional [7], many students dislike the differential scoring system that incentivizes them to be critical towards their predecessor. Even though PrESTO offers the possibility to appeal if students disagree with their received score, much stronger emotions can be seen using differential scoring than when using absolute scoring. Although we did not collect data on the students' individual learning styles, this is potentially an explaining factor [8].

In a follow-up evaluation in 2018-2019 (Fig. 5) we asked students how instructive they found the relay activities, and also how uncomfortable they felt while performing them. Students indicated that they learnt most from doing the primary task, and found being assessed the least instructive. Students felt most uncomfortable while assessing, followed by being assessed. These results about instructiveness and uncomfortableness suggest that students felt they learnt the most in their comfort zone, namely the primary tasks, analyzing a variety of cases and making the analysis consistent. For the tasks they felt most uncomfortable with (being assessed, assessing a review, assessing, and receiving feedback) they also indicated they learnt the least. Conceivably, these tasks were so far from their comfort zone that they entered the panic zone [9]. Reviewing, improving the predecessor's work, and receiving an appeal decision were experienced as uncomfortable, but also as instructive. During these activities, the students were probably outside their comfort zone, but not in the panic zone.

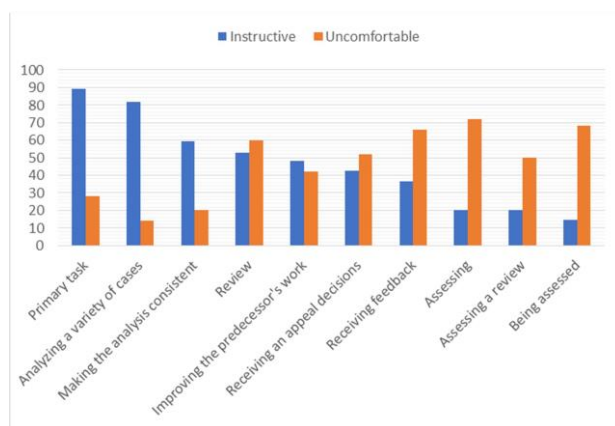


Fig.5. Instructiveness and the feeling of uncomfortableness of the learning activities during PrESTO

The incentives in the relay system have been specifically designed to stimulate students to do good quality work. However, care should be taken that students do not feel too uncomfortable while working on a relay. We think we can make it less uncomfortable for the students by enhancing their capacity for evaluative judgment, i.e., their capability to judge the work of oneself and others [10]. In line with [11], we plan to do this by providing better instruction in class, and by further improving our guidelines for formulating constructive feedback.

IV. CONCLUSIONS AND FUTURE WORK

The PrESTO system allows students to develop skills through repeated application by conducting assignments and assessing and improving work of others. The PrESTO software, which is available under an open source license at <https://github.com/pwgbots/presto>, enables students to move through the sequence of activities, and enables teachers to manage the workflows, even for very large class sizes. The system can also be used as a flexible peer review system by defining a relay of only one step, rather than a series of steps. We are presently developing <http://wiki.presto.tudelft.nl> to provide more information for instructors.

PrESTO has been used in a variety of courses over the past six years. This way of working can in theory be used for any course having open-ended assignments. Evaluations

show that students learn a lot, but also that a project relay is time consuming and can be quite stressful for them. Students indicated that they feel most uncomfortable while assessing and while being assessed. At the same time, they found these assessment-related tasks to be the least instructive.

To redress the emotional discomfort related to PrESTO without compromising the quality of learning, we plan to improve the way in which we prepare students for the review and assessment tasks. We will do this by providing students with better instruction and guidance for reviewing and assessing, and by highlighting the importance of the assessment tasks as part of their preparation for professional life. This preparation lies outside of the PrESTO system itself, but it should be part of the way of working in which the system is embedded. Within the system itself, we will work on specific incentives for students to write constructive reviews that are helpful to the students who are being assessed.

ACKNOWLEDGEMENT

This research was partially supported through the incentive scheme for open and online education of the Dutch Ministry of Education, Culture and Science.

REFERENCES

- [1] C.M. Brooks and J.L. Ammons, "Free Riding in Group Projects and the Effects of Timing, Frequency, and Specificity of Criteria in Peer Assessments", *Journal of Education for Business*, 2003, vol. 78, no. 5, pp. 268-272, doi: 10.1080/08832320309598613.
- [2] I. Kollar and F. Fischer, "Peer assessment as collaborative learning: a cognitive perspective", *Learning and Instruction*, 2010, vol. 20, no. 4, pp. 344-348, doi: 10.1016/j.learninstruc.2009.08.005.
- [3] K. Topping, "Self and peer assessment in school and university: reliability, validity and utility", in *Optimising New Modes of Assessment: In Search of Qualities and Standards*, M. Segers, F. Dochy and E. Cascallar, Eds. Dordrecht: Springer, 2003, pp. 55-87, doi: 10.1007/0-306-48125-1_4.
- [4] 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", *Computers in Human Behavior*, 2016, vol. 64, pp. 94-107, doi: 10.1016/j.chb.2016.06.020.
- [5] J. Yuan and C. Kim, "The effects of autonomy support on student engagement in peer assessment", *Educational Technology Research and Development*, 2018, vol. 66, no. 1, pp. 25-52, <https://doi.org/10.1007/s11423-017-9538-x>
- [6] H. Purchase and J. Hamer, "Perspectives on peer-review: eight years of Aropä", *Assessment & Evaluation in Higher Education*, 2018, vol. 43, no. 3, pp. 473-487, doi: 10.1080/02602938.2017.1359819.
- [7] N.K.L. Pope, "The impact of stress in self- and peer assessment", *Assessment & Evaluation in Higher Education*, 2005, vol. 30, no. 1, pp. 51-63, doi:10.1080/0260293042003243896.
- [8] P. Vickerman, "Student perspectives on formative peer assessment: an attempt to deepen learning?" *Assessment & Evaluation in Higher Education*, 2009, vol. 34, no. 2, pp. 221-230, doi: 10.1080/02602930801955986.
- [9] M. Brown, "Comfort zone: Model or metaphor?" *Journal of Outdoor and Environmental Education*, 2008, vol. 12, no. 1, pp. 3-12, doi: 10.1007/BF03401019
- [10] J. Tai, R. Ajjawi, D. Boud, P. Dawson and E. Panadero, "Developing evaluative judgement: enabling students to make decisions about the quality of work", *Higher Education*, 2018, vol. 76, no. 3, pp. 467-481, doi:10.1007/s10734-017-0220-3
- [11] A. Landry, S. Jacobs and G. Newton, "Effective use of peer assessment in a graduate level writing assignment: a case study", *International Journal of Higher Education*, 2015, vol. 4, no. 1, pp. 38-51, doi: 10.5430/ijhe.v4n1p38