Learning Analytics for Peer-assessment: (Dis)advantages, Reliability and Implementation

Blaženka Divjak   *blazenka.divjak@foi.hr*
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, Varaždin, Croatia

Marcel Maretić   *marcel.maretic@foi.hr*
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, Varaždin, Croatia

Abstract

Learning analytics deals with the data that occurs from students' interaction with ICT: collecting data, analyzing and reporting that can influence learning and teaching. Analysis of validity and reliability of assessment lags behind other applications of learning analytics. Mathematical modeling of learning analytics for assessment, especially for peer-assessment, is presented in the paper. In addition, students' recognition of advantages and disadvantages of peer-assessment is analyzed and categorized. Finally, implementations of reliability check of peer-assessment in Moodle Workshop module are explained.

Keywords: Learning analytics, assessment, peer-assessment, metrics, reliability and validity of peer-assessment

1. Introduction

Society today is characterized by a rapid social and economic change. From accelerating evolution of ICT arise needs for new competencies such as self-regulated and peer learning, evaluation of peer work and metacognitive skills. The usual critique toward online tasks is that they rarely meet the requirements for development of higher order skills and higher order knowledge. Entwistle states that "*Some of these advances [in e-learning], however, have done little more that move information around in more efficient*

ways." (cf. [11], p. 138). Their development is enabled by deep learning (cf. [10]) and assessment has a clear connection with learning outcomes (cf. [1]) that comprise key competencies. Our research is based on the *Embedded Assessment Paradigm* (cf. [23]), where learning analytics are used in order to interpret data about students' learning, to assess their academic progress, to predict future performance and to personalize educational process. The 2015 edition of the Horizon report learning analytics [14] as a midterm trend in education on a 3–5 year horizon, and just a year later in Horizon report 2016 learning analytics and adaptive learning along with *Bring Your Own Device* (BYOD) are expected to be increasingly adopted by higher education institutions in one year time [15].

We have conducted action research during the three year period in the course *Project Management* at the Master Level of Entrepreneurship study programme at the Faculty of Organization and Informatics (FOI) at University of Zagreb in which 107 students were enrolled in academic year 2013/2014 and 2014/2015. Assessment and learning tasks were carefully prepared in the blended learning environment and clearly connected with intended learning outcomes of the course and the study programme (cf. [4]). Preliminary research has been published in [6].

This paper is organized as follows. After briefing the current state of the art, we investigate the possibilities of combining peer-assessment with learning analytics to enhance deeper learning approach by students. Specifically, we propose a new metric to measure reliability of peer-assessment and self-assessment. For initial prototyping we present the results on the data gathered in the last three years of the *Project Management* course. Finally, based on literature, students' feedback and authors own experience, we analyze advantages and disadvantages of peer-assessment.

2.   Learning Analytics for Assessment: State of the Art

Learning analytics (LA) as a research field is quite new but already very propulsive and influential. Still, the research arena is just shaping and its research methods are still under construction. LA analyzes data primarily, but not exclusively produced by student's interactions with information and communication technology (ICT) and especially with Learning Management System (LMS) where huge quantity of data is stored. The least contested definition of learning analytics is "*Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimizing, learning and the environment in which it occurs.*" This definition, according to [12], originated at the first international Conference on Learning Analytics and Knowledge (LAK2011) and was adopted by the Society for Learning Analytics Research (SoLAR) [17].

As an interdisciplinary field LA is positioned at the intersection of business intelligence, web analytics, educational data mining and recommender/recommendation

systems (cf. [12]). Application area of LA is certainly in formal and informal education but also in non-formal learning. LA is basically all about learning. Gašević and Dawson in [13] stress: "*That is, instructors expressed their preferences of learning analytics features that offer insights into learning processes and identify student gaps in understanding over simple performance measures. With such insights, instructors can identify weak points in the learning activities performed by their students; topics the students have struggled with, and provide instructive and process related feedback on how to improve their learning.*" Ellis and Ferguson in [9] discuss definition of *Learning Analytics and Knowledge* and point out two limitations: (1) *limited usefulness* from both practical and pedagogical perspective; and (2) *limited focus* where only a portion of the student body is considered with too often students that are neither at risk nor the best forming an "overlooked middle". Further, the author argues that in "...*the scholarship on learning analytics, assessment data are almost never considered or referred to as part of the available data sets that can inform learning analytics.*" The reason behind this, she argues, is most likely "...*a direct product of the fact that, until relatively recently, the possibility of collecting and collating assessment data at a level of granularity that is meaningful and useful has simply been unthinkable.*" Finally, among several sets of assessment data, [9] mentions "*achievement mapped against explicit learning outcomes or assessment criteria (e.g., rubrics results)*".

This paper argues for the need and opportunity of utilizing results from granular assessment criteria (rubrics) in order to have insights into students learning as well as to evaluate the reliability and validity of student peer-assessment. Our motivations and research ideas come from education practice and mathematics, specifically geometry and metric spaces.

We use e-assessment embedded in LMS *Moodle* for assessment of complex problems and authentic tasks (cf. [4, 5]). A shift from computer-based assessment towards embedded assessment is happening in the area of e-assessment (cf. [23]).

3. Peer-Assessment

There is a considerable pool of research on employability and 21st century skills. Skills identified by these researchers as most wanted and important for long-term employability are the ability for managing own learning and peer-earning, the ability to successfully work in groups, ability of objective judgment of peer-work (critical thinking) as well as metacognitive skill of reflection about own learning and performance [22]. Consequently, we should strive to enhance and develop exactly these skills through formal, informal and non-formal learning.

Formative assessment and feedback help students take control of their own learning, i.e. become self-regulated learners ([21]). According to [24], peer-assessment and self-assessment offer the following four advantages:

(1) *Logistical* – because it saves teachers time.

(2) *Pedagogical* – because judging the other students work is an additional opportunity for students to deepen their understanding about a topic.

(3) *Metacognitive* – because grading can help to demystify testing and students become more aware of their own strengths, progress and gaps in knowledge and skills.

(4) *Affective* – because these types of assessment can make students more productive and cooperative, and thus can build a greater sense of shared ownership for the learning process.

Peer-assessment engages students to become more active learners, take responsibility for their learning, to apply deeper learning strategies and to gain a better understanding of their own subjectivity and judgment. At the same time, we (see [6]) recognize several disadvantages of peer-assessment which we classify in the following four groups and add possible strategies to mitigate them based on the literature review and our own experience.

(1) *Logistical* – because students need additional briefing time and teacher has to plan extra time for discussion of assessment criteria, goals, write some instructions in LMS, implement scoring rubrics etc.

(2) *Reliability risk* – because students are assessing their own peers. Some of their peers can be their friends and others can be members of other cliques in the classroom. Therefore teacher must be aware of it and if necessary anonymize assessment tasks.

(3) *Equalizing* – because of a tendency to award everyone the same mark. Learning analytics can aid the discovery of assessment patterns (especially in large groups).

(4) *Metacognitive* – because not all students are well equipped to undertake peer-assessment and they have not developed metacognitive skills so far. Therefore, teacher should start with the self assessment tasks that have lower stakes to train the students and use LA analysis to analyze reliability of peer-assessment whenever necessary (big groups, high stakes assessments).

Finally, students' peer-assessment can only be considered a satisfactory substitute for teacher assessment if the grading results are very similar to the teachers' assessment. If students' grades are not reliable, the teacher must override the assessment [24]. Further, we must be aware that peer-assessment of simple tasks (for example determining whether a claim is correct) is much easier than grading a complex task such as an essay, a problem solving task or a project. In the later case students must be guided in their assessment tasks by discussing and explaining grading criteria and their weights (cf. [5]).

3.1. Final grade calculation

Final grade for a received set of peer-assessments can be calculated in several different ways. The most common option is to calculate the average of received grades for each criteria. For example, column-wise mean for each criteria on the data in Table 1 yields a tuple

$$(2.66, \ 1.66, \ 2, \ 0.66, \ 1.0)$$

as a final grade for this set of peer-assessments. Other options for calculation of the final grade are discussed in [7].

	C_1	C_2	C_3	C_4	C_5
peer$_1$	3	2	2	0	1
peer$_2$	2	2	2	1	1
peer$_3$	2	1	2	1	1

Table 1. Example data for scoring rubric peer-assessment

The maximal amount of available points for criteria C_i are very important and have to be determined with care by teachers and/or by a more sophisticated method. Author in [5] demonstrated an approach based on the multi-criteria group-decision making with groups formed by representatives of several stakeholders (teachers, students, former students, employers, etc.).

3.2. Case Study

LMSs offer packages for integration of self-assessment, peer-assessment and summative assessment. These packages support some automatic analysis of learner data. In our case study on a *Project Management* (PM) course we used the Moodle package *Workshop* (see [20]) for assessment support and data collection. In the course peer-assessment has been used for two assessment activities: peer-assessment for essay writing (low stake assessment) and peer-assessment of projects (high stake assessment). The first peer-assessment is used as a preparation for the assessment of projects that accounts for 30% of the final grade. Scoring rubrics were used for peer-assessment of both activities.

We analyzed students' comments on peer assessment gathered in academic years 2013/2014 and 2014/2015. In the academic year 2014/2015 of the PM course we have analyzed peer-assessment data at the criteria for two peer-assessment activities: essays and projects. The goal here was to analyze and compare the reliability of peer-assessment at summative and criteria level.

4.   Validity and Reliability of Peer-Assessment

In this section we address the question of validity and reliability of peer-assessment.

> "*Assessment is **valid** if it has to measure what was intended ... Assessment is **reliable** if an equivalent grading would be given if marked again shortly afterwards or by another person. If assessment is not reliable, it cannot be valid; but an assessment can be reliable and yet be invalid, by accurately measuring the wrong thing.*" (Entwistle in [11], p. 157).

Validity and assurance of assessment is quite hard. Preparation of the teaching, learning and assessment with the use of constructive alignment is the first step in this process. Validity of assessment is evaluated relative to the intended learning outcomes of study programme and consequently the course. Alignment of assessment with the learning outcomes (LOs) can be prepared in a variety of ways. Besides learning outcomes, the type of assessment must be chosen to take into account students' prior knowledge, the size of a class, teacher's workload, available resources etc. Additional options for verification of validity of the assessment can be performed through the use of student questionnaire on the achievement of LOs and through continued tracking of students in their later career.

Reliability is studied through analysis of variations that occurs across raters and across different peer-assessment of the same individual rater. The former is commonly referred to as *intra-rater reliability*, while the latter is known as *intra-rater reliability* [16].

Summative reliability, as analysis of the span of received totals, for the dataset of the *Project Management* course was analyzed in [6] and suggested that about 85% of assignments received sufficiently reliable peer-assessments. However, we argue that summative reliability is inadequate as it is lacks the necessary to provide valuable feedback to students. Therefore, by utilizing only summative reliability, an opportunity for deeper learning may be missed.

4.1.   Reliability Measure

Detailed look of peer-assessments analyzed on the criteria level is interesting and reveals more information about reliability of peer-assessment.

Let $S = (c_1, c_2, \ldots, c_n)$ and $S' = (c'_1, c'_2, \ldots, c'_n)$ be tuples which represent gradings S and S' of the same essay according to the criteria $C_1, C_2, \ldots, C_n$. Gradings S and S' can be imagined as points in n-dimensional space.

The common and naïve approach is the use of Euclidean metric as a distance measure. Instead, we propose the use of the normalized 1-metric (known as taxicab of

Manhattan distance, cf. [2]) in n-dimensional space, where n is the number of criteria in the scoring rubric.

Distance between points S and S' can be calculated as normalized Manhattan distance scaled down to $[0, 1]$ range:

$$d(S, S') = \frac{1}{N} \left(|c_1 - c_1'| + \cdots + |c_n - c_n'| \right) \, ,$$

where N is a total (maximum) number of available points for this activity. Normalized taxicab distance is a scaled sum of absolute differences by criteria.

Let $\mathcal{S}$ be a set of peer-assessments for the same assignment. We base the modeling of the reliability on the diameter of the peer-assessment set, i.e. a maximal pairwise distance between gradings in $\mathcal{S}$:

$$\max_{S, S' \in \mathcal{S}} d(S, S') \, .$$

A divergent set of assessments $\mathcal{S}$ has a larger diameter, whereas a small diameter is expected for a coherent set of assessment. We propose the reliability measure $\mathrm{rel}(\mathcal{S})$ of the grading set $\mathcal{S}$ as

$$\mathrm{rel}(\mathcal{S}) = 1 - \mathrm{diam}\, \mathcal{S} \, .$$

Reliability of the grading set takes values in the $[0, 1]$ range with 1 being the maximal reliability of the perfectly uniform grading set.

We consider two options for gradings sets that are not acceptable, i.e. insufficiently reliable, and propose adequate solutions.

1. In case when low reliability is a result of diverging set of assessment the solution is to demand supervision, i.e. to ask for teacher's grading of the assignment and in that case teacher's grade becomes the final grade. This approach is feasible in classroom of school settings with manageable number of students.

2. As low reliability of a grading set can be a result of an outlier grading (a singular odd grading, quite different that the rest), it is possible to exclude the outlier grading and consider the reliability of the remaining set.

   This idea has application in situations with a sufficient number of available peer-assessments (MOOC setting) with the potential to decrease the teacher's workload.

4.2.   On the choice of metric and criteria of assessment

The use of the taxicab distance metric also known as Manhattan or rectilinear metric is advocated for several reasons. Distance between gradings can be calculated without the need for paper or a calculator. Also, a grade change on a single criterium exactly equals the total change. Figuratively, we may say that a grade "moves" in a rectangular grid, just like the Manhattan taxicab. Finally, the total of points, as the value most interesting to the student, is exactly the Manhattan norm of the grading vector. As taxicab metric *is* a metric in a mathematical sense, i.e. it satisfies the axioms of a metric, it has no disadvantages when compared with Euclidean metric.

  Normalization, realized as scaling of the metric to the $[0, 1]$ range, is introduced to allow for comparison of results and to facilitate recommendations in forthcoming implementations. For example, to allow a recommendation to take a grading set with a relative diameter less than 0.3 as acceptable.

4.3.   Granular vs. summative reliability

Peer-assessments are commonly analyzed only at the summative level, i.e. by measuring differences of sums (totals) of received assessments from different raters. Tentative comparison of summative and granular reliability of received peer-assessments for our datasets suggests that a significant proportion of about 10 to 15% of assessment sets seem reliable at the summative level, but are not reliable when looked at the criteria level. In these assessment sets raters show significantly greater variety at the criteria level than at the summative (total) level in rating these assignments. This is a missed opportunity for deeper learning when raters would receive inadequate feedback for these assessments.

  As mentioned before, our case study was performed on *Project Management* course in academic year 2014/2015 and on peer-assessment of student essays. Total of 62 students participated in the peer-assessment. Calculated Pearson's correlation of diameter and summative difference of assessment sets is $r = 0.59$ for the PM course's "essay activity" dataset and $r = 0.66$ for the PM course's "projects activity" dataset. This indicates that granular reliability is not a trivial extension of summative analysis.

5.   Students' Perception about Peer-Assessment

Finally, by utilizing the learning analytics collected in LMS we answer the following research questions:

  1. What is student perception about peer-assessment, assessment standards and criteria and mutual learning activity?

2. Is deeper learning encouraged by peer-assessment?

Students' views on peer-assessment were collected through closed questions in questionnaire and by open questions in a form of e-journal in LMS. Students' questionnaire was filled out by 45 students out of 62 for the academic year 2014/2015. The question relevant for peer-assessment was asked in the form of agreement with the claim: "*Peer assessment of essay and projects motivated me on new way of thinking and learning.*"

It follows that 73.33% of students agree or even strongly agree with the claim that peer-assessment and mutual learning is motivating and that it opened new ways of learning for them. More details in [6].

Furthermore students commented the peer-assessment exercise in a free form journal in a period of two consecutive academic years.

Students' perspective on whether deeper learning was encouraged through peer-assessment was taken in the form of the e-journal where students answered the following four questions:

1. *What you have learned through peer learning?*

2. *Do you see link of peer learning to course learning outcomes?*

3. *Was peer learning interesting?*

4. *How to enhance the peer learning exercise?*

Additional students' comments were welcomed.

Students' comments are systematized and presented in Table 2 to support advantages and disadvantages recognized in the literature.

We confirm the existing groupings and additionally recognized one advantage (*Empowering*) and one disadvantage (*Accountability*) based on the received feedback. Namely, students commented that peer-assessment is very useful and enhances the development of their own skill set through assessment capability. Some students, at the same time, feel uncomfortable with the responsibilities that arise in peer-assessment. They consider themselves incompetent for assessment and consider it a teacher's job.

Pros		Cons	
Logistical	– *It depends on the size of the group and level of their metacognitive skills* – *Monitoring of the process is also demanding*	Logistical	– *What is the meaning of criteria and levels? I find that teachers should teach us the the best practice for adequate assessment.*

Table 2 – continued from previous page

Pros			Cons	
Pedagogical	– *This was an opportunity to learn something new related to the theme of the assignment we had to assess. Careful listening to the oral presentation of the assignment is not enough to fully comprehend the material.* – *Assessment of asignments is more interesting and meaningful than mere listening of oral presentations taking 20 or 30 minutes.* – *Analysing others is beneficial for my own writing regarding style, ideas etc.*	Reliability		– *I find it hard to remain objective when I'm assessing the work of someone very close to me.* – *Assessment could be enhanced (better objectivity, easier to do) if we included more criteria.*
Metacognitive	– *We'll learn the most after tearcher grades these assignments. Then we will be able to see where we stand – what was correct, and what was missed, what must be improved in order to be objective in assessment. Assessment is hard.* – *Now we see what is valued and where emphasis is put in assessment.*	Equalizing		Observed by teacher (span of grades given by students vs span of grades given by teachers)
Affective	– *It was awkward to criticise and comment my colleagues, but I applied the skills learned in other courses, i.e. to shape a critique as advice.* – *It is a creative activity, it encourages out-of-the-box thinking* – *I learned through my own mistakes and realized that my assessments are too "generous".* – *I assume a different role when I question my knowledge. It was stressful in the beggining.* – *This activity has a positive effect on student's cognitive skills. We have not tried this before.*	Metacognitive		– *I think that most assessments are based on superficial impressions because they lack the time and/or ambition for deeper analysis.* – *This was my first assessment of peer's work.* – *Could be enhanced/improved with a quiz activity to check what was learned and memorized.* – *Interesting, should be introduced before final year.* – *Had to think out-of-the-box to grasp the significance of this type of activity. As we haven't done this before, it is not easy.*

Table 2 – continued from previous page

Pros		Cons
Empowering	– *Assessment of peer's assignment is very useful and closely related to the learning outcomes of this course.* – *Peer-assessment of essays and projects was hard and stressful (especially for projects). Nevertheless, I find it was helpful in reflection of my own skillset and assessment capability.*	Accountability — – *I don't want to take responsibility for the grades of other students.* – *We (the students) don't have the necessary competencies for assessment.* – *I think that teachers must read all of the assignments and have a final say.* – *I wish we tried this on a neutral example.* – *I assessed the work of my friends, not my employees. Assessment is teacher's job, not students'!*

Table 2. Assortment of representative students' comments gathered in a period of two consecutive academic years 2014/2015 and 2015/2016.

6. Implementation

A support for transparent and meaningful peer-assessment learning analytics is lacking in assessment analytics in general. Reliability check of assessment is the core issue assessment and especially peer-assessment. We analyze the current implementation in the Moodle LMS that is Open Source, used at FOI.

6.1. Moodle Workshop plug-in

Peer-assessment activity in the Moodle LMS is provided by the Moodle Workshop module. In peer-assessment activity, students receive two grades: a grade for their work and a grade for the quality of their assessments of other student's assignments.

Each participant in workshop is graded first for his submission, and later for his peer-assessment(s). These grades are visible as separate grade items in student's gradebook.

Calculation of the assignment grade for student's submission in Workshop is fairly simple. It is calculated as weighted mean of all received assessment grades without any reliability check. If the teacher wishes to influence the final submission grade, she/he can provide her assessment and set its weight (to a perhaps high value). The teacher can also entirely override received peer-assessments and set the final grade of the submission. The implemented logic behind the grade for assessment is more complex: it tries to estimate the quality of each assessment. This functionality is available through a subplugin "Comparison with the best assessment". In order to estimate the quality of

assessment, this subplugin selects the best assessment and rewards other assessments according to their distance from the best assessment.

Due to constraints of existing calculations and omission of reliability check we are developing a new Workshop subplugin that addresses the problem of grade calculation in peer-assessment, offers alternative methods for final grade calculation and most importantly offers the analysis of the reliability of peer-assessment at the criteria level. This subplugin implements the ideas and methods presented in this paper as well as in [7, 8].

7.   Conclusion

Assessment guides learning and therefore it has to be carefully prepared, conducted, analyzed and continuously improved. Especially important issues of assessments that need to be addressed are validity and reliability. In the case study of the PM course presented in [6] it was shown that peer-assessment can be constructed to be valid and reliable. Here we analyze reliability issues for peer-assessment where a difference can be observed in case of using granular approach based on assessment criteria instead of th total score.

So far, our findings agree with previous related research (cf. [24]) and confirm the inadequacy of summative analysis of reliability. We propose the use of scaled Manhattan metric based on taxicab norm to model both inter-rater and intra-rater reliability.

Further research should be directed toward investigation of proposed metrics for evaluation of peer-assessment for complex tasks such as a problem solving task of a project-based learning.

Students perceived the value of peer-assessment as they reported that peer-assessment combined with peer-learning is motivating; opens new learning paths and triggers deeper learning. Based on the literature and received students' feedback we support previously recognized advantages and disadvantages of peer-assessments. We recognize additional advantage and disadvantage category. Finally, there are the following systematization of advantages: *Logistical*, *Pedagogical*, *Metacognitive*, *Affective* and *Empowering*. Disadvantages are categorized as *Logistical*, *Reliability*, *Equalizing*, *Metacognitive* and *Accountability*.

Acknowledgement

References

[1] Biggs, J. "Aligning teaching and assessing to course objectives", in *Assessment*, vol. 19, no. 2, 2003, pp. 13–17.

[2] Divjak B., "Notes on Taxicab Geometry", Scientific and Professional Information Journal of Croatian Society for Constructive Geometry and Computer Graphics (KoG), 5, 2000, pp. 5–9.

[3] Divjak B., "Teaching methods for international R&D project management", in *International Journal of Project Management*, 26, 3, 2008, pp. 258-267.

[4] Divjak B., "Implementation of Learning Outcomes in Mathematics for Non-Mathematics Major by Using E-Learning", in *Teaching Mathematics Online: Emergent Technologies and Methodologies*, (Editors: A. A. Juan, M. A. Huertas, S. Trenholm, and C. Steegmann), IGI Global, 2012, pp. 119–140.

[5] Divjak B., "*Assessment of Complex, Non-Structured Mathematical Problems*", in *IMA International Conference on Barriers and Enablers to Learning Maths*, (Editors: Hersh, M.A., Kotecha, M.), 2015.

[6] Divjak B., Maretić M., "*Learning Analytics for e-Assessment: The State of the Art and One Case Study*", Central European Conference on Information and Intelligent Systems, Conference Proceedings, (Editors: Hunjak, T., Kirinić, V., Konecki, M.), 2015, pp. 81–87.

[7] Divjak, B., Maretić, M., "Geometry for Learning Analytics", Scientific and Professional Information Journal of Croatian Society for Constructive Geometry and Computer Graphics (KoG), 2016. pp. 48–56.

[8] Divjak, B., Grabar, D., Maretić, M., Assessment analytics for peer-assessment: a model and implementation, in *Learning analytics for curriculum and program quality improvement (PCLA 2016)*, CEUR Workshop Proceedings, 2016. pp. 27–31.

[9] Ellis, C., "Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics", in *British Journal of Educational Technology*, vol. 44, no. 4, 2013, pp. 662–664.

[10] Entwistle, N. J., "Approaches to studying and perceptions of university teaching-learning environments: concepts, inventory design and preliminary findings", in *Powerful learning environments: Unraveling basic components*, 2003, pp. 89–108.

[11] Entwistle, N. J., "Teaching for understanding at university: deep approaches and distinctive ways of thinking". Basingstoke, Hampshire, Palgrave Macmillan, 2009.

[12] Ferguson, R., "The state of learning analytics in 2012: a review and future challenges", in *Tech. Rep.* KMI-12-01, vol. 4, no. March, 2012, p. 18.

[13] Gašević, D., Dawson, S., Siemens, G., "Let's not forget: Learning analytics are about learning", in *TechTrends*, vol. 59, no. 1, 2015.

[14] Johnson, L., Adams Becker, S., Estrada V., and Freeman A., "NMC Horizon report: 2015 Higher Education Edition", The New Media Consortium, Austin, Texas, 2015.

[15] Johnson, L., Adams Becker, S., Cummins M., Estrada V., Freeman, A. and Hall, C., "NMC Horizon Report: 2016 Higher Education Edition", Austin, Texas, The New Media Consortium, 2016.

[16] Jonnson, A., Svigby, G., "The use of scoring rubrics: Reliability, validity and educational consequences", in *Educational Research Review*, 2007.

[17] LAK11, 1st International Conference on Learning Analytics and Knowledge, https://tekri.athabascau.ca/analytics/, accessed on 10.05.2016.

[18] Moodle LMS https://moodle.org/, accessed on 10.05.2016.

[19] Moodle LMS Plugins, available at https://moodle.org/plugins/, accessed on 10.05.2016.

[20] Moodle Workshop Plugin, Documentation available at

https://docs.moodle.org/29/en/Using_Workshop, accessed on 10.05.2016.

[21] Nicol, D. J., Macfarlane-Dick, D., "Formative assessment and self-regulated learning: a model and seven principles of good feedback practice", in *Studies in Higher Education*, Vol. 31(2), 2006, pp. 199–218.

[22] OECD "Skills Outlook 2015 – Youth, Skills and Employability", OECD Publishing, http://dx.doi.org/10.1787/9789264234178-en, accessed on 10.05.2016.

[23] Redecker, C., Johannessen, Ø., "Changing Assessment - Towards a New Assessment Paradigm Using ICT", in *Eur. J. Educ.*, vol. 48, no. 1, 2013, pp. 79–96.

[24] Sadler, P., Good, E., "*The impact of self-and peer grading on student learning*", in *Educ. Assess.*, vol. 11, no. 1, 2006, pp. 37–41.