

# Learning Analytics for e-Assessment: The State of the Art and One Case Study

Blaženka Divjak and Marcel Maretić

Faculty of Organization and Informatics

University of Zagreb

Pavlinska 3, 42000 Varaždin, Croatia

{blazenka.divjak, marcel.maretic}@foi.hr

**Abstract.** *Learning analytics deals with the data that occurs from students' interaction with ICT: collecting data, analyzing and reporting can influence learning and teaching. Application of learning analytics for analyzing assessment has lagged behind other areas of application. We argue here for the need for learning analytics for assessment with a special focus on peer and self-assessment. All forms of assessment should encourage deep learning. Reliability and validity of peer assessment will be discussed. In this context, a case study will be presented.*

**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 competences 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 than move information around in more efficient ways.” (cf. [8], p. 138). Their development is enabled by deep learning (cf. [7]) and assessment has a clear connection with learning outcomes (cf. [1]) that comprise key competences. Our research is based on the *Embedded Assessment Paradigm* (cf. [13]), 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 reports learning analytics [11] as a midterm trend in education on a 3–5 year horizon.

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 Informatics (FOI) at University of Zagreb in which 131 students were enrolled. 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]).

After briefing the current state of the art, we investigate the possibilities of combining peer assessment with learning analytics to enhance deeper learning.

Specifically, we propose metric to measure peer assessment and self-assessment reliability and discuss validity of peer and self-assessment.

For initial prototyping we present the results on the test data gathered in the last three years of the *Project Management* course.

## 2 Learning Analytics for Assessment: State of the Art

Learning analytics (LA) as a research field is quite new. The research arena is just shaping and its research methods are still under construction. Learning analytics deals with analysis of data 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 following definition of learning analytics is the least contested: “*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 [9], originated at the first international Conference on Learning Analytics and Knowledge (LAK2011) and was adopted by the Society for Learning Analytics Research (SoLAR).

LA is, as an interdisciplinary field, positioned at the intersection of several disciplines: business intelligence, web analytics, educational data mining and recommender/recommendation systems (cf. [9]). LA's motivations and research ideas come from education science and mathematics, specifically geometry and metric spaces. Application area of LA is certainly in formal and informal education but also in non-formal learning. Basically LA is all about learning. Gašević and Dawson in [10] stress: “*That is, instructors expressed their pref-*

Table 1: Constructive alignment on the PM course

Study program LOs – relevant for the course	Course LOs related to the study programme LOs	Teaching and learning method	Assessment method	Student workload – ECTS credits
Apply concepts and methods of project management ... Explain and critically evaluate project cycles as well as project success criteria	Write a project application Analyze and evaluate project success criteria Present a project to a professional audience ...	Students work individually on essay writing and peer-evaluation	Teacher assessment, self-assessment and peer assessment based on prepared criteria and scoring rubrics	15 – 20h = 0.6ECTS (approx. 15% of the course's 4 ECTS)
		Students work in teams on task of project application writing and self- and peer-evaluation		30 – 40h = 1.5ECTS (approx. 30% of the course's 4 ECTS)

ferences 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.” Further, Ellis and Ferguson in [6] 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, [6] 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.

At the same time we are using e-assessment embedded in Moodle Learning Management System (LMS). It is possible to implement e-assessment for complex problems and authentic tasks (cf. [5]). In the area of e-assessment shift has been made from computer-based assessment towards embedded assessment (cf. [13]).

In our approach we are less inclined to conform to the paradigm of Explicit Testing. We are much closer to the Embedded Assessment paradigm which does away with tests and instead, via Learning Analytics, uses the data produced during the learning process as a basis for providing feedback and guidance to both learners and teachers (see [13]).

### 3 Assessment and Peer Assessment

Skills most wanted and most important for long-term employability are ability for lifelong and peer learning, to successfully work in groups, making judgments about peer work as well as metacognitive skill of reflecting on her/his own learning and performance. Consequently, we should strive to enhance and develop exactly these skills through formal and informal learning.

Formative assessment and feedback can help students take control of their own learning, i.e. become self-regulated learners ([12]). According to [14], peer assessment and self-assessment have 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.

In general this means that students are more active learners, more responsible for their learning, apply deeper learning strategies and have a better understanding of their own subjectivity and judgment. At the same time, we (the authors) recognize some possible disadvantages of peer assessment which we classify in the following four groups:

- (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* i.e. tendency to award everyone the same mark. Learning analytics can help, especially with bigger groups, to discover assessment patterns.
- (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 comparable to the teachers'. If students' grades are not reliable, the teacher must override the assessment [14]. Further, we must be aware that peer assessment of simple tasks (determining whether a claim is correct) is much easier than grading a complex task such as essay, problem solving 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]).

Assessment packages for LMSs have been developed to integrate self-assessment, peer assessment and summative assessment. These packages also often integrate the automatic analysis of learner data. In our case study we have used a package *Workshop*<sup>1</sup> in the Moodle LMS as assessment support and data collection. Students are able to submit their work during the *Workshop* activity. Submissions can be assessed by teachers, self-assessed, or assessed by peers (students). The *Workshop* also allows multi-criteria assessment based on scoring rubrics. Students can obtain two grades in a single *Workshop* activity – one grade for their submission (that is how good their submitted work is) and another grade for their assessment (that is how well they assessed their peers).

## 4 Case Study: *Project Management Course*

We have conducted action research in a period of three years in the course *Project Management* (PM) at the Master Level of Entrepreneurship in which 131 students were enrolled. 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 (details in [4]).

Constructive alignment (cf. [1]) has been prepared to pair learning outcomes (LOs) of the study program with the course LOs, and also to connect course LOs with teaching and learning methods, assessment tasks and

student workload. The problems of specific LOs of PM were described in [3].

For the research presented in this paper we considered only LOs relevant for the peer assessment. The constructive alignment for two LOs of this study programme is presented in the Table 1. Careful preparation of constructive alignment is essential for validity of assessment.

Table 2 lists assessment tasks at PM course along with their percentage value relative to the total course grade.

Table 2: Assessment tasks at PM course

Assessment task	Percentage of the total mark
Tasks in the classroom and online LMS (Moodle)	17
Essay (writing + peer assessment + artifact + journal evaluation)	9+3+2+1 = 15
Tasks in computer labs	8
Project application writing and presentation	30
Tests (2)	30
<b>Total</b>	<b>100</b>

First two generations of students (2012/13 and 2013/2014) had their tasks assessed only by teachers based on well-defined assessment criteria and rubrics. An innovative way of mutual learning and peer assessment based on those same rubrics has been created for the last generation of students (2014/15) where *students assessed themselves*.

There were two tasks where peer assessment was used: essay grading and project grading. The first task (essay grading) with smaller stakes was also used to train and prepare students to assess according to criteria and rubrics in LMS Moodle and enhance student understanding of assessment standards and criteria.

In the essay grading we used the following criteria  $C_i$  (with total weight  $r_i$  of the criteria  $C_i$  is listed in parenthesis):

$C_1$ . Topic covering, soundness ( $r_1 = 3$ );

$C_2$ . Essay structure ( $r_2 = 2$ );

$C_3$ . Text formatting, pictures, graphs, examples ( $r_3 = 2$ );

$C_4$ . Language and grammar ( $r_4 = 1$ );

$C_5$ . Referencing ( $r_5 = 1$ ).

Criteria and levels were described in detail in the scoring rubrics. Everything was implemented in the Moodle *Workshop* package for assessment. Table 3 describes phases in essay writing, peer assessment and peer learning. Help for students for each activity has been provided in the LMS or in the classroom (as is indicated in Table 3). Students chose a topic for essay in the LMS and then had two weeks to prepare an essay due to the instructions,

<sup>1</sup>[https://docs.moodle.org/19/en/Workshop\\_module](https://docs.moodle.org/19/en/Workshop_module)

recommendations and scoring rubrics provided for them. After submittal of the essay, the second phase began – the peer assessment.

Each student got three essays randomly assigned to her/him by the LMS for assessment. Peer assessment was performed with the scoring rubrics. Written feedbacks were also required. To enhance mutual peer learning group work (3-4 students in a group) followed. Students that had similar essay topic worked together and needed to summarize the main accents from their topics in the form of presented artifacts (not Power Point, 5 minute duration). Later task, the grading of the project as a new round of peer assessment, was prepared by taking into account students' comments following the peer assessment of the essays.

Finally, by utilizing the learning analytics collected in LMS in this period (three years) we can answer the following research questions:

1. How to prepare peer assessment to be reliable and valid and at the same time enhance mutual learning?
2. What is student perception about peer assessment, assessment standards and criteria and mutual learning activity?
3. Is deeper learning encouraged by peer assessment?

## 5 Validity and Reliability of Peer Assessment

In this section we try to answer the first research question regarding 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.”* (see [8], p. 157).

Checking and assuring validity of assessment is a hard problem. 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 with the intended learning outcomes of study programme and consequently the course. Correspondence of assessment with the LOs can be prepared in many ways. Besides LOs, the type of assessment depends on the students' prior knowledge, the size of a class, teacher's workload, available resources etc. One possible assessment structure for the PM course is given in Table 1 and Table 2. Additional options for verification of validity of the assessment can be performed through the use of student questionnaire querying about achievement of LOs and through tracking students in their career. Students' perspective on the results of peer assessment will be presented in the next section.

Our aim in the case study is the analysis of the peer assessment of essays. Activities, type of students work, available help and duration are presented in Table 3.

Reliability for peer assessment for PM was checked by comparing the gradings from academic year 2014/2015 ( $n = 62$  students) with two previous academic years ( $n = 34 + 35$  students) when only teachers graded essays according to the criteria (scoring rubrics). As can be seen from Table 4, the results are comparable. These results correspond with the research in [14].

Comparison of assessment results during three years provides a starting point in analysis of assessment reliability (cf. [8]).

For a measure of reliability we have considered the span of totals of peer assessments of the same work: peer gradings whose span is within 2 points (i.e. less than or equal) are considered reliable (consistent); peer gradings that exceed 2-point span indicate inconsistent gradings and such gradings are flagged as unreliable (requiring supervision).

Table 5 presents some data on reliability at the overall grade level and suggests that students' evaluations are sufficiently reliable.

It is even more interesting to compare grading on the criteria level. For such analysis we have to introduce appropriate metrics. The common and naïve approach is the use of Euclidean metric. Instead, we propose the use of the normalized 1-metric (known as taxicab or Manhattan distance, cf. [2]) in  $n$ -dimensional space, where  $n$  is the number of criteria in the rubrics. Let  $S = (c_1, c_2, \dots, c_n)$  and  $S' = (c'_1, c'_2, \dots, c'_n)$  be tuples describing two student gradings  $S$  and  $S'$  of the same essay according to the criteria  $C_1, C_2, \dots, C_n$ .  $S$  and  $S'$  can be imagined as points in  $n$ -dimensional space. Distance between points  $S$  and  $S'$  can be calculated as normalized Manhattan distance:

$$d(S, S') = \frac{1}{n} \left( \frac{|c_1 - c'_1|}{r_1} + \dots + \frac{|c_n - c'_n|}{r_n} \right),$$

where  $r_i$  is a weight of the criterium  $C_i$ .

Let  $\mathcal{S}$  be a set of peer assessments (for the same work). As a measure for divergence of the assessment set  $\mathcal{S}$  we propose taking maximal pairwise distance between points in  $\mathcal{S}$ :

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

Manhattan distance (based on taxi-cab norm) is used because of the discrete nature of the assessment data. Normalization is introduced to allow future comparison with calculations based on different metrics.

Weight  $r_i$  of criteria  $C_i$  can be determined by teachers and/or by group decision making with the use of multi-criteria decision making. A group in decision making is usually heterogeneous and consists of representatives of teachers, students and other stakeholders (former students, employers etc., cf. [5]).

Table 3: Phases in essay writing and assessment

Activity	Type of work	Help	Duration	Result
Short essay writing	Individual	Instructions about good essay writing given before activity	2 weeks	Essay submitted on online (Moodle)
Peer assessment of essays	Each student assess three essays	Assessment rubrics with criteria and instructions provided online (in Moodle)	1 week	Peer review feedback online + comments on what student assessor learned
Mutual learning	Group work 3-4 with connected topics	Help provided in classroom, but it can be also by webinar	1 day	Key messages and interesting findings in form of working hints
Artifact – summary building	Group work 3-4 with connected topics	Recommendation with links provided online (in Moodle)	2 weeks	Artifact that summarized topics submitted online – 5' presentation in the classroom
Evaluation of activity and learning	Individual diary online	A few general questions given in the form of e-journal (in Moodle)	1 day	Evaluation reports online

Table 4: Comparison of assessments by academic year

Academic year	Average
2014/15 ( $n = 62$ )	6.40/9
2013/14 ( $n = 34$ )	(5.62/10) = 6.24/9
2012/13 ( $n = 35$ )	(6.11/10)=6.79/9

Table 5: First check of reliability

Number of grades within the 2-point span	Number of grades that equal or exceed the 2-point span
51	11
85%	15%

## 6 Students' Perception about Peer Assessment

Answers on the second and the third research questions are based on students' perception. There are two principal ways how students' views on peer assessment were collected: through closed questions in questionnaire and by open questions in a form of e-journal. 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."

The results are presented in Figure 1. 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.

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

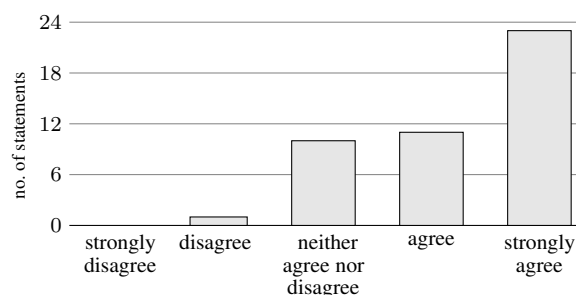


Figure 1: Agreement with: "Peer assessment of essay and projects motivated me and introduced me to a new way of thinking and learning." ( $n = 45$  answers)

- What you have learned through peer learning?
- Do you see link of peer learning to course learning outcomes?
- Was peer learning interesting?
- How to enhance the peer learning exercise?

Additional comments were welcomed.

Most common comments on learning and importance were:

- Interesting and important (both – the course and the peer assessment)
- I benefited from reflection on my own work – I had to see where I was not so good and I had to spot my own errors
- I learned from others how to better structure an essay and how to do it in a more interesting way
- We have learned more from assessment than in student's presentation of Power Point slides
- To learn how to assess is not easy, especially when you perform criteria-based assessment

- *I was surprised how objective an assessment can be, even when the essays were assessed by us (the students).*
- *This is a good preparation for assessing a real project.*
- *This is important for future professional work – encourages concise and structured writing, quick reporting and assessment based on defined criteria*
- *I was taught to respect various approaches and opinions when supported by arguments*
- *I appreciate the link between theory and practice*
- *I found out that assessing essays in a short period of time is hard – now I have much more respect for teachers' work*

Certain useful suggestions from students were implemented in the second peer assessment task in the course. Students suggested that:

- *More recommendation on structure should be given;*
- *Criteria should be explained in detail and to introduce more criteria and subcriteria;*
- *students dislike binary criteria (language, references);*
- *More time in classroom should be dedicated to discuss how to write assess it and what results should look like;*
- *Assessment should be anonymized.*

## 7 Conclusion

Assessment guides learning and therefore it has to be carefully prepared, conducted, analyzed and enhanced. Especially important characteristics of assessments are their validity and reliability. For validity analysis it is important to introduce constructive alignment with intended learning outcomes but also to take students' perspective on their achievements as well as to track their careers after graduation. For peer assessment reliability has to be carefully checked because several disadvantages can challenge reliability of results. We propose the modified Manhattan metrics (based on taxi-cab norm) to be used in order to check on reliability and further develop in the scope of learning analytics. Future research in the interdisciplinary field of learning analytics of assessment that include modeling by different metrics arising from non-Euclidean geometry or multi criteria decision making is needed. In the case study of the PM course presented in the paper it was shown that peer assessment can be constructed to be valid and reliable. Further, student perception is that peer assessment together with peer learning is motivating and opens new learning paths and that it trigger deeper learning approach. In that respect

further research should be done specially in the peer assessment of more complex tasks such as problem solving or project tasks.

We are acutely aware of the limitations of our current research – data is limited (since it is gathered from a single course). Therefore, it is too early for generalization. So far, our results agree with previous related research (cf. [14]). Further research should be directed toward investigation of appropriate metrics for evaluation of peer assessment (especially for peer assessment of complex tasks such as problem solving, projects, etc.) and new pedagogical applications of learning analytics.

## References

- [1] Biggs, J. "Aligning teaching and assessing to course objectives", *Assessment*, vol. 19, no. 2, pp. 13–17, 2003.
- [2] Divjak, B. "Notes on Taxicab Geometry", *Scientific and Professional Information Journal of Croatian Society for Constructive Geometry and Computer Graphics (KoG)*, 5, pp.5-9, 2000.
- [3] Divjak, B., Kukec, S. "Teaching methods for international R&D project management", *International Journal of Project Management*, 26, 3, pp. 258-267, 2008.
- [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*, A. A. Juan, M. A. Huertas, S. Trenholm, and C. Steegmann, Eds. 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*, 2015.
- [6] Ellis, C. "Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics", *Br. J. Educ. Technol.*, vol. 44, no. 4, pp. 662–664, 2013.
- [7] Entwistle, N. J. "Approaches to studying and perceptions of university teaching-learning environments: concepts, inventory design and preliminary findings", in *Powerful learning environments: Unravelling basic components*, 2003, pp. 89–108.
- [8] Entwistle, N. J. "Teaching for understanding at university: deep approaches and distinctive ways of thinking". Basingstoke, Hampshire: Palgrave Macmillan, 2009.
- [9] Ferguson, R. "The state of learning analytics in 2012: a review and future challenges", *Tech. Rep. KMI-12-01*, vol. 4, no. March, p. 18, 2012.

- [10] Gašević, D., Dawson, S., Siemens, G. “*Let’s not forget: Learning analytics are about learning*”, TechTrends, vol. 59, no. 1, 2015.
- [11] Johnson, L., Adams, S., Estrada, V., Freeman, A., “*NMC Horizon Report: 2015 Higher Education Edition*”, Austin, Texas, 2015.
- [12] Nicol, D.J., Macfarlane-Dick, D., “*Formative assessment and selfregulated learning: a model and seven principles of good feedback practice*”, no. June 2015, pp. 37–41, 2006.
- [13] Redecker, C., Johannessen, Ø., “*Changing Assessment - Towards a New Assessment Paradigm Using ICT*”, Eur. J. Educ., vol. 48, no. 1, pp. 79–96, 2013.
- [14] Sadler, P., Good, E., “*The impact of self-and peer grading on student learning*”, Educ. Assess., vol. 11, no. 1, pp. 37–41, 2006.