

## **Learning analytics approach of EMMA project**

Kairit Tammets  
Centre for educational technology  
Tallinn University  
Estonia  
Contact – Kairit@tlu.ee

Francis Brouns  
Welten Institute  
Open Universiteit, Nederland  
Contact - Francis.Brouns@ou.nl

MOOCs have been criticized for narrowing the learning scope, removing a human element from the processes of teaching and learning, replacing reflection from the student learning process, manipulating the knowledge they present and consequently limiting the connection with social and cultural aspects of knowledge acquisition (Cooper, 2013). Pedagogical value of the MOOCs can be doubtful due its teacher-centered approach and with the considerable focus on content. Bates (2012) addresses the myth that xMOOCs are a new pedagogy and notes that so far the teaching methods in MOOCs are based on old and out-dated behaviorist pedagogy, which rely primarily on information transmission, computer-marked assignments and peer assessment. Also the business model behind the MOOCs has been questionable. But still several educational institutions (Stanford University in particular) and researchers have discovered that MOOCs provide excellent possibilities to study the learning in online settings and MOOCs can be a medium for educational research, especially through the implementation of learning analytics (LA). Coffrin, Corrin, de Barba et al (2014) claim that although there are many challenges providing MOOC participants with high quality educational experience at such a large scale, MOOCs have created significant opportunities for researchers and educators who are interested in learning analytics. Learning environments and platforms track and log users' traces, which form massive data sets of valuable learning experiences with the aim to analyze different processes and patterns and can be used for predicting the learning behavior. Results can be used in design and development of traditional learning management systems or virtual learning environments used by higher education or K12 institutions. It means that algorithms developed for LA need data to be tested and often the testing will be performed with the existing data sets from some other (MOOC) platform or system. Key to the usefulness of learning analytics is the ability to provide data to educators in a way and format that it supports their decision-making about educational interventions and curriculum design (Coffrin et al, 2014).

EMMA platform is developed by the European Commission 7<sup>th</sup> FP project. EMMA delivers free, open, online courses in multiple languages from different European universities. EMMA will operate in two main modes; as an aggregator and hosting system of courses produced by European universities; and as a system that enables learners to construct their own learning pathways using units from MOOCs as building blocks. Learning analytics features will be designed and developed for EMMA platform. Currently the work is still in process, but in the autumn 2014, when the first MOOCs will be piloted, learning analytics application and methodology will be tested and evaluated.

Learning analytics in EMMA project will focus on: a) real-time analytics through learning analytics dashboards for instructors and students; b) retrospective analysis of the digital traces in EMMA platform. More broadly the learning analytics approach of EMMA is built on the idea of Greller and Drachsler (2012) that the potential of learning analytics to become a powerful tool for informing educators and supporting students, providing a platform for better understanding and predicting student performance and learning needs. Still one of the characteristics of EMMA platform is the integration of a video transcription and translation system. As a result, seven transcription systems and eight translation systems are provided. Another important feature of EMMA platform is providing a personal learning environment (PLE) that allows learners to pick and mix from those sections of the various MOOCs on offer that they consider important and relevant and build their own personalized courses. Therefore one of the main aims is to investigate if the automatic translation and customization module has an impact on learners' progress and enrollment, are there more enrolments because MOOCs are offered in multiple languages and does it lead to more students actually starting with the MOOC, or if it results in more students completing the MOOC. Another aim is to support the personal learning paths and personal learning needs of the EMMA MOOC learners. That again puts additional requirements and restrictions to the LA we want to design

Additionally in the pilot phase of the MOOCs in EMMA platform, the following analysis will be performed:

- Clustering of the participants - In the pilot phase, MOOC participants of EMMA platform will be clustered. Clusters will be in the pilot phase probably: a) enrolled b) not started (enrolled, but have not returned to course); c) lurker; d) passive; e) active; f) drop-ins – enrolls, but is active in one-two weeks only. There is a chance that in EMMA participants can be at the same time “active” and “drop-ins”. As EMMA is a MOOC platform then calculating the dropout (after excluding the lurkers and drop-ins), although it is acknowledged in the project that calculating drop-out rate is not efficient for making any conclusion about MOOCs or their participants. Also as the EMMA approach supports the participants to combine their own courses from different building blocks of different MOOCs, then calculating the completion rates is not the primary interest of EMMA. Clustering in general enables to get insights about participants' behavior, which is needed for evaluating the courses and analyze for example in what stage the participants become less active. Pilot phase evaluation demonstrates in which way it is possible or needed to cluster the participants.
- Progress and performance - Students' progress in EMMA MOOCs is measured in accordance with the course lessons. Each course consists of lessons that consist of several units. Units consist of the materials, videos and assignments. Progress will be visualized in the students' dashboards. Students can see their efforts during the whole course or within a module.
- Uptake of knowledge - In EMMA the learning analytics approach for the uptake of knowledge is mainly related with the social structures and has to consider available functionalities of EMMA. The uptake is mainly analyzed based on the interactions in the conversation functionality (later also weblog and comments could be added to the analysis).

- Social structures - social network and artefacts analysis will be performed in order to find out what kind of social structures emerge in the MOOC context and in which way the materials and resources mediate between the participants and what kind of networks emerge around the materials or artefacts.
- Engagement with the content - for supporting the evaluation of the course design, access and use of the learning materials will be evaluated. The intensity and frequency of the accessed learning materials will be provided to the MOOC instructors

The technical architecture of the EMMA learning analytics application is illustrated in Figure 1. Different components developed by several partners constitute the learning analytics application: tracking system, learning record store (LRS) Learning Locker for storing the tracked events, dashboards for MOOC participants and MOOC instructors.

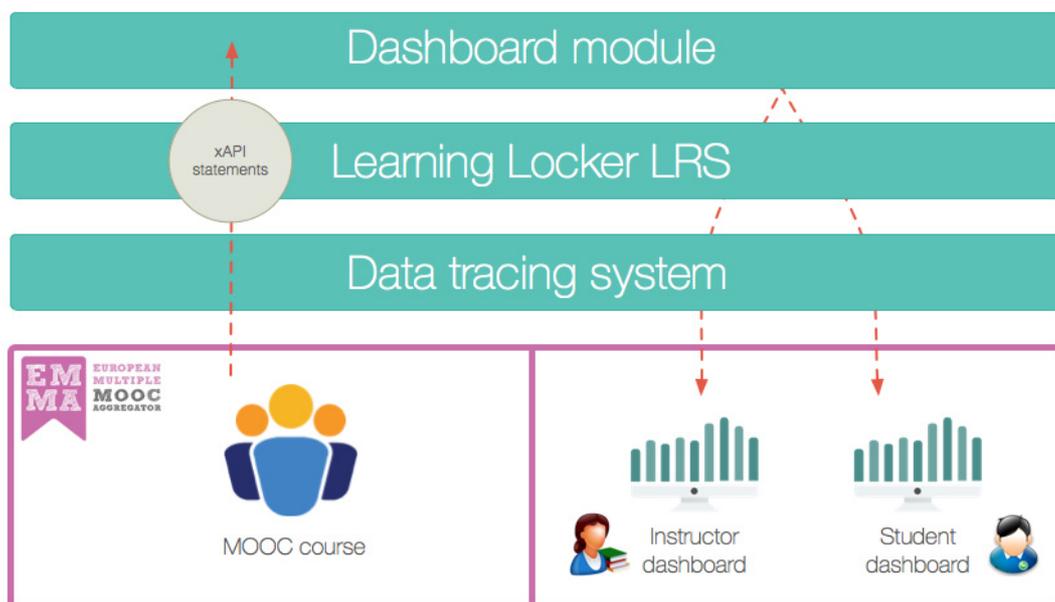


Figure 1: EMMA Learning analytics application (EMMA project deliverable 4.1, 2014).

One of the main interoperability specifications that is used in the learning analytics domain and also in EMMA for analyzing the learning is the Experience API<sup>1</sup> or xAPI. The xAPI is based on the idea of tracking activity through Activity Streams, which was developed to provide a better way of expressing social media activity than existed before (Cooper, 2014). The EMMA tracking tool exchanges the xAPI format with the LRS, which is able to exchange data with the final application. In the case of EMMA, Learning Locker is the LRS of choice, as it is the reference open-source LRS. Although Learning Locker provides a dashboard to visualize data to participants, in the case of EMMA dashboards will be developed from scratch.

The focus of the EMMA learning analytics dashboard will be to support awareness and sense-making of the learning activities in online settings. Coffrin and colleagues

<sup>1</sup> <http://www.adlnet.gov/tla/experience-api/en> <http://www.tincanapi.com>

(2014) have pointed that current MOOC platforms are limited in their ability to provide data feeds and visualizations that can be used to assist instructors, students and others in their decision making. They claim that the true value of learning analytics to MOOCs will be realized when platforms embed the analyses and visualization for end users to routinely and easily access and use. In the pilot phase, the EMMA instructors' dashboard visualizes: students' progress during the lesson or the course as a whole and lessons' overview. Weekly the snapshot of social interactions will be visualized in the course context based on posts and replies in discussion section and comments in blog (visualized through SNA). Students' dashboard visualizes social interactions in the course context based and progress in the course context based on the performed self-assessment tasks and assignments, accessed materials and time spent on them, participation in discussion and number of activities completed. Also EMMA aims to provide recommender dashboards. The learner has a chance to choose from different modules in different MOOCs and to create their personal learning path for supporting personalized bottom-up approach to learning. For supporting students to find suitable MOOCs and select needed modules for their learning path, their dashboard will provide meaningful recommendations.

In EMMA data set will be created, but the main aim is not to share the sets with the third parties, therefore the data sharing aspect is not explicitly formulated in the project. It is still unsure how suitable datasets from other types of platforms are to develop and test LA algorithms, in particular how generalizable those are to other contexts especially when not all learning environments support the same pedagogies. That might influence the data that can potentially be tracked, traced and stored. For MOOCs the traditional pedagogies can't be applied due to the massive number of students, which means that conventional type of data might not be available. Moreover, challenges in MOOCs are different, and LA should focus on different aspects than are important in regular educational models. That again dictates and also limits the kind of data that becomes available and can be tracked and stored. It also dictates the interpretation, because students and staff have specific needs attuned to that particular pedagogy. We believe that MOOC can't be designed according to conventional educational paradigm, because it is not sure how interchangeable datasets are from other learning environments.

EMMA tracking tool will collect quite a lot of users' data because as more than 15 MOOCs will be implemented in EMMA platform during the project. Collected data is mainly used for improving the LA application, because in the initial development phase there is no data used and algorithms are not tested. Algorithms currently base on assumptions what kind of data and variables are needed to track and it is not clear yet if the outcome of the algorithm is valuable, feasible, informative. It means that there is need to make some an educated guess' and several pilots to test and evaluate them.

Therefore the data set is particular for MOOCs, and in particular for the five aspects that EMMA wants to evaluate (from pedagogy to instructional design; from web learning platform to MOOC-based platform; from linguistic isolation to linguistic inclusiveness; from country-specific platform to a pan-European platform (language, culture); from individual MOOCs to integrated learning). Due its specific analysis aims it might not be a good data set to exchange for other LA purposes.

In case the data sets created in EMMA would be created with the purpose to exchange sets with other researchers, platform developers, MOOC providers, the data set needs to be completely anonymised to ensure privacy and confidentiality of users' data. Just removing the usernames from the data is not enough. Data set may contain user written content like in forum or blogs, it is likely that user enters names or other details that may identify them. By combining different variables that are being tracked it is quite often possible to identify individuals depending on what kind of data is being tracked.

In EMMA we apply LA foremost to support learners and teachers. Of course we will track and store personal data. There is no way to get around that in a learning environment where we have to support the learner. However, we will inform the users that a LA application is running, that we track certain data, that those data will be analysed to support them, that no personal data is being shown to other learners but will be shown to teachers, that in general visualisations shown to other learners is based on aggregated data. EMMA will take precautions to ensure data is stored securely and safely and can't be accessed by others. No data will be knowingly made available to other people outside the project. LA data will be used as well to evaluate the algorithms and the platform. Users will be informed about this. Data for evaluation purposes will only be made available to those researchers from partner organisations that are responsible for the evaluation.

### **Acknowledgements**

This work is funded by the EU, under the Competitiveness and Innovation Framework Program 2007-2017 (CIP) in the European Multiple MOOC Aggregator (EMMA) project. Grant Agreement no. 621030.

### **References**

- Bates, T (2012). What's right and what's wrong about Coursera-style MOOCs? <http://www.tonybates.ca/2012/08/05/whats-right-and-whats-wrong-about-coursera-style-moocs/>
- Coffrin C., Corrin, L.E., De Barba, P.G., & Kennedy, G.E. (2014). Visualizing patterns of student engagement and performance in MOOCs. In Pistilli M, Willis J, Koch D, Arnold K (eds), *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*. ACM Press. Teasley S, Pardo A.
- Cooper, A. (2014). Learning Analytics Interoperability – The Big Picture in Brief. An introductory briefing. Retrieved from <http://laceproject.eu/publications/briefing-01.pdf>
- Cooper, S. (2013). MOOCs: Disrupting the University or business as usual? Retrieved from <http://arena.org.au>
- EMMA project. (2014). D4.1 Learning analytics: theoretical background, methodology and expected results.
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, 15(3), 42-57