Learning Analytics, the thorny issue of data gathering for MOOCs

Ilaria MERCIAI

Università degli Studi di Napoli Federico II, Napoli (NA)

Abstract

One of the critical issues emerging from the scientific debate about MOOCs is related to what approach to learning analytics (LA) to adopt. Within the context of MOOCs, in fact, the issues araising from the floor concern the main purposes and challenges underlying tracking and data analysis as well as tools and methodology that best serve the "teaching by data" purpose. This paper seeks to relate our experience of trying to adapt and extend LA debate and tools to the Federica system, since the Federica Web Learning portal of the University Federico II, (http://www.federica.unina.it/) is now launching MOOC courses, complete with timeframe and assignments.

Keywords: MOOC, Learning Analytics, improve learner experience

Introduction

Learning Analytics (LA) is defined as the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs". This definition dates back to the 1st International Conference on Learning Analytics and Knowledge (LAK) in 2011.

In 2013, the *Horizon Report* described it as a "field associated with deciphering trends and patterns from educational big data, […] to further the advancement of a personalized, supportive system of higher education" and there appeared to be growing interest in the subject on the part of academic institutions.

These two strands of analytics, focusing on the quality of the content and the personalisation of the learning experience, are also mentioned by Greller and Drachsler (2012): «Learning Analytics will become a powerful means to inform and support learners, teachers and their institutions in better understanding and predicting personal learning needs and performance». However, they also intimate at differences between theory and practice, pointing out that «the processes and requirements behind the beneficial application of Learning and Knowledge Analytics, as well as the consequences for learning and teaching, are still far from being understood» (p.42). Clow also concluded that there is no substantial reference in the literature to application of LA to MOOCs (Clow, 2013).

The jury appears to be out on whether effective tools are yet available: «Analytics tools and techniques that focus on the social pedagogical aspect of learning are required» (Siemens, 2012, p. 3).

This formed the backdrop to the decision-making process regarding LA at Federica, the Web Learning portal of the University Federico II (Campus Virtuale Project, FESR 2007-2013), (http://www.federica.unina.it/) which is now preparing its MOOCs launch.

Learning data and learner classification

In the field of LA, at least two different levels of data collection, comprising quantitative and qualitative data, seemed to be important:

1) analysis of user behaviour: including an overview of traffic, enrolments per course, number of lessons followed and completed, assignments corrected, number and length of contributions to forum, hours of study, videos watched;
2) analysis of users and levels of user satisfaction: including qualitative feedback, engagements in participation as well as social media activities

Both levels overlap and form part of a third level which we called "functional" i.e. the basic data to enable teachers to monitor student presence, engagement and performance in case of awarding certificates of completion, or performance badges. This requires the provision of teacher as well as student dashboards.

While recognising the importance of the functional data, the Federica team wanted to go further, and investigate aspects of learning outcomes and sustainability.

Identifying learner type

Early research into MOOCs tended to highlight the high drop-out rates on this kind of course compared to more traditional education (Clow 2013) but the utility of drop-out rates as a measure of success is questionable when many MOOC users have no intention of finishing the course. The work of Dietz-Uhler & Hurn (2013) illustrated that from hearing about a course and deciding to register to actual knowledge acquisition is a natural funnel effect (Fig. 1), so far fewer learners actively interact with course content than start the course.

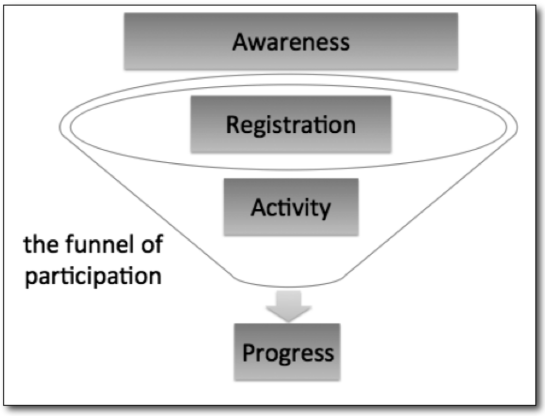

Figure 1: The funnel of participation. Source: Dietz-Uhler, B., Hurn J. E. (2013). *Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective*. Journal of Interactive Online Learning. Volume 12, Number 1.

This led certain authors like Milligan, Littlejohn & Margaryan (2013) to establish categories for MOOC students according to their behaviour online. These categories enable MOOC providers to differentiate the kind of LA they use, modifying it according to the specific category:

1) **enrolled** (explicitly "enrolled" the course);
2) **not started** (enrolled, but have not returned to course);
3) **lurker** (enrolled and has returned to the course once);
4) **passive** (enrolled and has accessed one material and participated in one discussion or submitted one assignment) (dissatisfied);
5) **active** (has accessed 50% of the materials and submitted 50% of the assignments or participated in 50% of the discussions);
6) **drop-ins** (enrolls, but is active in one-two weeks only and are satisfied).

This classification led to another round of questions: what information should we get from all users? How can we find out why the "not starters" didn't start? At what stage do we ask a drop-in or a passive learner to complete a questionnaire so it is early enough to have their cooperation but late enough to elicit useful feedback?

Brief focus on questionnaires

For information-gathering via questionnaires the prime consideration is not to overload learners, especially in the early stages and to use multiple choice questions to focus closely on strategic objectives regarding users and their response to both the platform and the MOOC courses. An initial possibility is to use questionnaires in 3 phases:

1) **Registration Questionnaire:** brief but including basic demographic data (age, gender)
2) **Expectation Questionnaire:** profiling (educational background, nationality, status, profession and qualifications) as well as supplementary information (native language); how they heard about Federica, reasons for enrolling, expectations from course;
3) **Exit Questionnaire:** whether expectations were met; reaction to specific aspects of courses e.g. format, videos, time-frame, assignments (number, type, level of difficulty).

Pierre Gorissen in an article in the Media and Learning Conference Newsletter (August 2014) highlighted how self-reported information in questionnaires is less reliable than tracking data.

This reinforced our intention to try and provide questions where learners did not feel they had to make a good impression or demonstrate their performance.

Final Remarks

The million dollar question still remains: what data is it strategically useful to collect?

Our analysis seemed to point at “levels and quality of student engagement and reasons for this”. The literature, as we stated above, focuses on the concept of improvement of content and services. In an ideal world “improve” would mean turning the data into information that leaders can use to make informed decisions” (Knowledge Advisors Report 2012). But how can we ensure that LA really “enhances the eLearning experience”? Pappas in a recent article gave “5 Reasons Why Learning Analytics are Important for eLearning” (predict learner performance; personalise learning experience; increase retention rates; improve future courses; boost cost efficiency) (Pappas 2014).

Our reflections left us with interesting food for thought. Exciting developments in technology allow for more sophisticated forms of data collection and analysis that could provide new insights into pedagogy and learning processes in general. But whilever the tracking devices remain platform specific, we can only measure what happens within the confines of any LMS, and ignore the learning that goes on outside that space. Basic Deweyan concepts of experience and learning through interaction are ignored. The thorny issue of how to extend tools to collect data encompassing the whole learning experience in a digital world remains.

References

Clow, D. (2013). *MOOCs and the funnel of participation.* Third Conference on Learning Analytics and Knowledge (LAK 2013), 8-12 April 2013, Leuven, Belgium.

Dietz-Uhler, B., Hurn J. E. (2013). *Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective*. Journal of Interactive Online Learning. Volume 12, Number 1.

Greller, W., Drachsler, H. (2012). *Translating Learning into Numbers: A Generic Framework for Learning Analytics*. Educational Technology & Society, 15(3), 42–57.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., and Ludgate, H. (2013). *NMC Horizon Report: 2013* Higher Education Edition Austin, Texas: The New Media Consortium.

Knowledge Advisors Research. (2012). *The Role and Function of a Learning Analytics Leader*. Trainingindustry.com

Larusson, J. A., White, B. (Eds.) (2014). *Learning Analytics, From Research to Practice*, Springer, New York: Springer.

Milligan, C., Littlejohn, A., & Margaryan, A. (2013). *Patterns of Engagement in Connectivist MOOCs*. Journal of Online Learning and Teaching, 9(2).

Pappas, C. (2014). *5 Reasons Why Learning Analytics are Important for eLearning*, Elearningindustry.com.

Siemens, G. (2012). *Learning analytics: envisioning a research discipline and a domain of practice.* LAK '12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. 4-8.

Acknowledgement

Special thanks to prof. Rosanna De Rosa and Ruth Kerr for our continual exchange of ideas on this and other topics.