The Use of Learning Analytics to Improve Online Learning Outcomes: A Systematic Literature Review

Rajabalee Yousra Banooor, Santally Mohammad Issack, Rennie Frank
Centre for Innovative and Lifelong Learning (CILL)
University of Mauritius (MAURITIUS)
University of the Highlands and Islands (UK)
y.rajabalee@uom.ac.mu, m.santally@uom.ac.mu, Frank.Rennie@uhi.ac.uk

Abstract

ICTs are transforming the bricks and mortar institutional delivery methods with virtual and online learning converging to such an extent that the learner can experience anytime, anywhere learning, irrespective of the mode of education that he or she is engaged into. Such learning transactions occur in environments which are extensively data-driven. Researchers have argued that understanding patterns in such wealth of available data could be of value to address drop-out issues in online learning, improve student engagement and performances as well as the overall learning experiences leading to better student satisfaction. This concept is referred to as Learning Analytics. In this paper, we conduct a systematic literature review to look at how learning analytics have been used to model learner engagement in online courses and how their engagement has influenced their performances.

We used Cooper’s taxonomy of literature reviews as research method. The method consists of five main steps: (1) identify the problem statement (2) search literature for data collection (3) evaluate the relevance of data (4) analyse and synthesize data (5) interpret and discuss the findings. In this study, the data collection process resulted in selecting 132 articles. This was followed by a first screening which identified only 77 articles for their pertinence in the field. The search was further refined to 40 articles for their appropriateness and relevance to address the aim of the study. We found that most learner engagement models in analytics studies used platform data and mainly access and activity logs of the system, and other data such as student participation in online discussion forums, and clicks on the platform to view pages, and do other types of tasks. We found that there is a gap in the research studies with respect to student engagement measurement, pertaining to courses designed as a set of activities which are outcomes and competency-based.

Introduction

Recently, the concept of learning analytics has been proposed as an effective approach to analyse the wealth of available information related to learner activity (Rienties et al., 2017; Eckerson, 2006; Johnson et al., 2011; Siemens et al., 2011). Learning analytics is defined as the measurement, collection, analysis and reporting of data about learners and their learning contexts, for the purposes of understanding and optimising learning and the environments in which it occurs (Verbert et al., 2012). The latter argued that learning analytics as an intelligent model, helps to advise on learner performance, seeks to find what influences their attention and perception, and then recommends suitable pedagogical resources to enhance and regulate their learning processes.

Buckingham and Ferguson (2012) and Siemens (2013) describe the concept of learning analytics as being the results of data analysed in relation to learner activity, which can be used to investigate the learners’ learning processes. Such information is perceived to be of help in predicting learner problems and preferences, and it is postulated that they can assist in the improvement of students’ learning experiences and promote student success rates in online courses (Olmos & Corrin, 2012; Smith, Lange, & Huston, 2012). For example, Smith et al. (2012) found that learners’ pace of learning and their engagement with their learning materials can be used to predict their performance and eventually create effective learning opportunities. It was found that learning design drives students’ learning experiences and have an impact on their outcomes and performances (Rienties and Toetenel, 2016; Rienties et al., 2018).

In studies on learner retention, researchers support the fact that frequently learners’ performance can rationally be predicted, in terms of social, psycho-emotional and demographic factors, by using learning analytics (Credé and Niehorster, 2012; Richardson, 2012; Marks et al., 2005). Similarly, Macfadyen and Dawson (2010) found that variables such as discussion forum posts and completed assignments, can be used as practical predictors of learner performance, and thus can be used to help in learners' retention and in improving their learning experiences. It
has been highlighted by researchers in the learning analytics area employed more human-led methods of discovery and retrieval, and then provided for holistic frameworks for actionable steps, such as guiding a tutor about ways a specific student is struggling, so that appropriate intervention can happen to facilitate the student’s learning process (Berland et al., 2014).

**Research Question and Methods**

To address the research problem of improving student’ online learning experiences using Learning Analytics both as a human-led model of discovery coupled with an intelligent component to advise on key actions to improve the overall learning design and intervention processes, a systematic literature review was conducted in line with the approaches proposed by Cooper (1988). Learning analytics (LA) is still an emerging research area but which has constantly been evolving from its most basic form of descriptive statistics to its close links to predictive models relying on machine learning and other intelligent techniques. The following research questions are set for this research.

**RQ 1:** How can Learning Analytics be used to model students’ engagement in online courses?

**RQ 2:** How can Learning Analytics be used to predict students’ performances from students’ engagement in online learning?

To address the research questions using the approach of Cooper (1988), the following steps were taken.

1. **Elaboration of a literature search strategy focusing on (i) students’ engagement in online courses, models of students’ engagement from the literature, (ii) other LA studies focusing on students’ online engagement and (iii) predictive analytics for RQ 2 with a focus on students’ engagement and performances.** Preferences for the LA studies were given to studies within the last 8 years while for online students’ engagement, the search covered the past decade. The selected databases for the literature review were the directory of open access journals, Science Direct, Emerald Insight, Google Scholar, amongst others.

2. **Only full papers in English language and peer-reviewed related to the main research questions were selected.** Papers were classified into the following categories: Empirical Research, Intervention Studies, Conceptual (Qualitative) Papers and Review Papers.

3. A summary of each paper including their methodology adopted and the main findings and discussion was done.

4. **Qualitative Analysis was carried out based on step 3 to conceptualize the relevance of learning analytics in the local context of the University of Mauritius.**

   Since the objective in this review was to collect relevant papers and identify what have been done in this context, with regards to the research questions, 132 articles were found. There have been a number of keywords that were used to search for relevant studies, for example: ‘learning analytics and student engagement’ OR ‘application of learning analytics in online course’ OR ‘student engagement in activity-based online learning approach’ OR ‘measure student performances with learning analytics’ OR ‘learning analytics and activity-based learning’ OR ‘learning analytics models for online learning’ OR ‘student engagement indicators for online learning’. The Figure 1 below, gives an indication of the criteria for identifying papers and it details the process for finalizing the selection of articles with respect to step 2 above. Initially, the literature search produced around 132 papers and following a first screening, only 77 papers were retained as the rest could not be used to specifically address the research questions of the study. Further scrutiny was undertaken to filter those papers which do not fall within the parameters of the aim of the study. As such, 43 articles were excluded in a first screening for being irrelevant to Higher Education and could not be used to address the research questions. 7 articles on MOOC were excluded as the context is different. 10 workshop papers, posters and abstract only were excluded. 5 articles on LA in on-campus courses were excluded. Exclude articles that dated more than 10-12 years ago. Books were excluded as full-text were not available.

   **Figure 1: Selection of studies**
a total of 40 papers were included for the full text review, after the validity and reliability analysis.

Observations

This section presents an analysis of the published studies and evaluates and interprets their findings by addressing the two research questions of this study.

RQ 1: How can Learning Analytics be used to model students’ engagement in online courses?

In order to address this research question, a classification of the relevant studies was done as shown in Table 1 below:

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Type of publication</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma, Han, Yang, and Cheng</td>
<td>2015</td>
<td>Research</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Henrie, Halverson, and Graham</td>
<td>2015</td>
<td>Review</td>
<td>Review</td>
</tr>
<tr>
<td>Viberg, Hatakka, Bälter and Mavroudini</td>
<td>2018</td>
<td>Review</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Florian, Glahn, Drachsler, Specht and Gesa</td>
<td>2011</td>
<td>Review</td>
<td>Actuator-Indicator model</td>
</tr>
<tr>
<td>Kotsiantis, Tseliou, Filippidi and Komis</td>
<td>2013</td>
<td>Research</td>
<td>Clustering and Visualization</td>
</tr>
<tr>
<td>Zotou, Tambouris, Triantafyllou, Timcenko, Busk Kofoed, Stracke, Riviou, García Barriocanal, Utz, Martos and Tarabanis</td>
<td>2017</td>
<td>Research</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Fallon, Walsh and Prendergast</td>
<td>2013</td>
<td>Research</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Hogaboam, Chen, Hmelø-Silver, Lajoie, Bodnar, Kazemitarab, Wiseman, and Chan</td>
<td>2016</td>
<td>Research</td>
<td>Visualization</td>
</tr>
<tr>
<td>Strang</td>
<td>2017</td>
<td>Research</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Shah, Nair and Richardson</td>
<td>2017</td>
<td>Review</td>
<td>Review</td>
</tr>
<tr>
<td>Khalil and Ebner</td>
<td>2016</td>
<td>Review</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Hampden-Thompson and Bennett</td>
<td>2013</td>
<td>Research</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Kagklis, Karatrantou, Tantoula, Panagiotakopoulos and Verykios</td>
<td>2015</td>
<td>Research</td>
<td>Text-mining</td>
</tr>
<tr>
<td>Gong, Liu and Zhao</td>
<td>2018</td>
<td>Research</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Lee, Song and Hong</td>
<td>2019</td>
<td>Research</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Richards</td>
<td>2011</td>
<td>Research</td>
<td>Analytics</td>
</tr>
<tr>
<td>Ahn, Butler, Alam and Webster</td>
<td>2013</td>
<td>Research</td>
<td>Descriptive analytics</td>
</tr>
</tbody>
</table>

Table 1: Relevant studies for Learning Analytics to model student’s engagement

The growing number of research studies carried out on learning analytics has been emphasized. Learning analytics is a way to report data left as digital footprints by learners in virtual learning environments, for example their activity, participation and contributions to a discussion forum can be analysed to improve the quality and value of their learning experience. Viberg et al., (2018) carried out a review of papers on LA research and its evidence for teaching and learning in Higher Education and they postulated that there is little evidence that learning analytics improves learning outcomes. However, there is evidence (35%) based on the analysis of studies that fall under the years 2012-2018, that it improves learning support and teaching in higher education. On the other hand, the study of Arnold and Pistilli, (2012) who used an application called Signals to allow tutors to identify disengaged students, reported that the use of learning analytics has a positive impact on academic performance with an increase in satisfactory grades and a decline in withdrawals. Similarly, Tempelaar et al., (2017) and Tabuenca et al., (2015) recorded an improvement in learning performance when using the practice of learning analytics. Therefore, we can assume that the analysis of data to identify any learning issues can be done using the capabilities of learning analytics, and it can potentially provide a learner engagement model to promote academic performance in the interest of the learners. Research on learners’ retention (Credé & Niehorster, 2012; Marks, Sibley, & Arbaugh, 2005; Richardson, 2012) have suggested that academic performance can be predicted when the findings from the learning analytics process are applied and thus, prediction can potentially help improve learner retention.
The study of Ma et al., (2015), used learning analytics to track data related teaching and learning activities in order to build an interaction-activity model to demonstrate how the instructor’s role has an impact on students’ engagement activities. According to them, the instructor plays an important role that positively influences students’ engagement in online learning. The instructor’s role is extended not only to that of a facilitator by guiding the process of learning but also by using a variety of methods to design effective learning activities that will engage students in learning. Their study revealed that when students are engaged in learning, they are encouraged to complete learning tasks assigned to them and this exerts a positive influence on their learning.

Moreover, online learning tasks which are stored on e-Learning platform can also provide evidence for competence assessment through the lens of learning analytics. Florian et al., (2011) supported this proposition and described how an activity-based learner model can be constructed based on aggregators of learner activities on Moodle by monitoring and observing the actions of the learner in order to assess his or her competences. Similarly, Rayón et al., (2014) used the SCALA (Scalable Competence Assessment through Learning Analytics) approach to extract learning metrics from students’ learning activities in order to discover learning patterns and assess parameters that could support competency assessment process. As such, the progress and engagement of the learner can be detected, while capturing evidence related to competency development. And, this gives teachers the possibility of taking corrective measures to adjust their teaching in order to support students’ learning and performances.

There are also studies which have integrated learning analytics in activity-based learning environment, as part of students’ learning experience and in order to gain an insight of their performances (Zotou et al., 2017). The activity-based learning approach help students to develop their knowledge, critical thinking, practical and collaborative skills through the completion of a series of ‘learn-by-doing’ activities. These activities can be in the form of concept mapping, written submission and brainstorming discussions (Fallon et al., 2013). The study of Fallon et al., (2013) used the NSSE (National Survey of Student Engagement) questionnaire to measure and report on students’ engagement in learning materials and activities. They found encouraging results whereby they could establish that students responded positively to the activity-based learning approach and there has been an enhancement in students’ participation and engagement. Robinson and Hullinger (2008) also argued that the NSSE is found to be an appropriate initiative to measure engagement in online courses as it provides useful results. The study of Richards (2011) also builds on similar techniques by using the NSSE questionnaire as learning analytics approach. However, it was found that there is no real indication that learners were engaged with their teachers’ postings and the depth of engagement in terms of intellectual value could not be determined.

On the other hand, Hogaboam et al., (2016) used visual dashboards to investigate teachers’ understanding of students’ learning activities in an activity-based online learning environment. The idea was to use a combination of students’ logs and their self-report data to understand and interpret their activities that could be useful to inform how LA dashboards should be designed. In this context, the design of dashboard can incorporate a range of useful features to allow teachers to identify student’s engagement and disengagement. However, the visualisation of data did not provide teachers with any understanding about students’ progress as the quantified indicator did not provide any meaning to the instructor. Therefore, it was important to find the correct combination of data that would help identify lurking students and track students’ progress.

Strang (2017) emphasizes on measuring the link between students’ graded outcomes and their online activities from a business course. It was found that when students are encouraged to complete online lesson such as quizzes, this promote their learning and engagement and hence result in higher grades. Hence, many studies (Shah et al., 2017; Khalil and Ebner, 2016; Gong et al., 2018) have outlined the use of learning analytics as a mean and a tool to discern patterns of behaviour and determine key performance indicators of learners’ progress in e-learning environments. Lee et al., (2019) reported on how the different indicators of student engagement, can help to improve student engagement and ultimately assist tutors in effective curriculum designs. These indicators are psychological motivation, peer collaboration, cognitive problem solving, interactions with instructors, community support, and learning management. They are related to activities where the learners are engaged in developing their critical thinking skills, building their knowledge, communicating and collaborating with peers and tutors.

Gong et al., (2018) concluded from their study that learning analytics interventions can improve students’ academic performances by intervening in their learning process to analyse and report on their data. Alternatively, in open courses, which target large scale of learners, descriptive analytics can be used to provide an understanding of learner participation and engagement (Ahn et al., 2013). Students’ engagement was also modelled by the analysis of their emotions through their participation in forums and their performance in online course (Kagklis et al., 2015). They found that students’ participation in forums is not directly associated with their performance. This is because at the beginning of the course, when students are faced with difficulties, they tend to communicate
with their peers and tutors more. However, later, it was observed that most students prefer to emphasize on working on their assignment as they will be given access to their exam, upon completion of a cumulative number of assignments and obtaining their grades. Therefore, although students tend to slow down their participation, they were still engaged in the online course.

**RQ 2: How can Learning Analytics be used to predict students’ performances from students’ engagement in online learning?**

In order to address this research question, a classification of the relevant studies was done as shown in Table 2 below:

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Type of publication</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietz-Uhler and Hurn</td>
<td>2013</td>
<td>Review</td>
<td>Review</td>
</tr>
<tr>
<td>Tempelaar, Heck, Cuypers, van der Kooij and de Vrie</td>
<td>2013</td>
<td>Research</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Tempelaar, Rienties, Mittelmeier and Nguyen</td>
<td>2018</td>
<td>Research</td>
<td>Clustering</td>
</tr>
<tr>
<td>Tabaa and Medouri</td>
<td>2013</td>
<td>Research</td>
<td>Data-integrator</td>
</tr>
<tr>
<td>Khalil and Ebner</td>
<td>2015</td>
<td>Review</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Sin and Muthu</td>
<td>2015</td>
<td>Review</td>
<td>Review</td>
</tr>
<tr>
<td>Gašević, Dawson, Rogers and Gasevic</td>
<td>2016</td>
<td>Research</td>
<td>Correlational</td>
</tr>
<tr>
<td>Strang</td>
<td>2017</td>
<td>Research</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Fallon, Walsh and Prendergast</td>
<td>2013</td>
<td>Research</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Hogaboam, Chen, Hmelo-Silver, Lajoie, Bodnar, Kazemitabar, Wiseman, and Chan</td>
<td>2016</td>
<td>Research</td>
<td>Visualization</td>
</tr>
<tr>
<td>Agudo-Peregrina, Iglesias-Pradas, Conde-González, and Hernández-García</td>
<td>2014</td>
<td>Research</td>
<td>Mixed-mode</td>
</tr>
<tr>
<td>Shum</td>
<td>2012</td>
<td>Review</td>
<td>Review</td>
</tr>
<tr>
<td>Scheffel, Drachsler, Stoyanov &amp; Specht</td>
<td>2014</td>
<td>Research</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Xing, Guo, Petakovic &amp; Goggins</td>
<td>2015</td>
<td>Research</td>
<td>Activity Theory</td>
</tr>
</tbody>
</table>

**Table 2: Relevant studies for Learning Analytics to predict engagement and performance**

Learning analytics can be used to create a predictive model to predict learners’ academic performance. Educators can follow and observe learners’ behaviour and interaction based on various combined data, identify any problems, then determine the potential outcomes. In this way, interventions and coaching, if needed, can be made in a timely manner to promote the success of learners. Tempelaar *et al.* (2013) used learning analytics through the measurement of learning related data such as demographic, entry test, learning dispositions, learning management system and quiz data to predict students’ performances in blended courses. On the other hand, Agudo-Peregrina *et al.*, (2014) did not find any consistent patterns in using learning analytics for prediction of performances in blended courses.

Tempelaar *et al.*, (2018) derived prediction models built on rich data in the form of learning dispositions combined with learning analytics trace data to identify students who are at-risk of failure. They found that providing students with formative feedback have strong predictive power to support learner to succeed. Therefore, the use of learning analytics not only inform on students’ performances but also how they can be used to improve retention and help students succeed (Dietz-Uhler and Hurn, 2013; Tabaa and Medouri, 2013; Khalil and Ebner, 2015).

Furthermore, Hussain *et al.*, (2018) focussed on developing a predictive analytic model, to determine the level of engagement of students in VLE learning activities. They postulated that when integrating such a model in a VLE, tutors would be allowed to identify low-engaged students and evaluate their engagement through different assessments. As such, it becomes easy for tutors in the effective design of their course materials for an increased level of engagement of students.

Moreover, Clow (2013) stressed on the fact that the theoretical basis of learning analytics, to predict experimental observations, is not clear and defined, although it is the study of educational data. For this reason, researchers (Gašević *et al.*, 2016; Clow, 2013; Dawson, 2008; Macfadyen & Dawson, 2010; Suthers *et al.* 2008) claimed that
the field of learning analytics is still an exception rather than a rule because it is significant to identify which relevant variables should be considered to represent learners’ effort. The choice of what is to be measured is important for accurate predictions (Gašević et al., 2016). There are many learning variables that actually contribute and impact on learners’ learning experiences, only some of which are visible, and only some of these which can be measured easily. In their relevant findings, Temperlaar et al. (2013) argued that many factors such as cultural differences, learning styles, learning motivations and learning emotions might be influential on learning. Hence, the debate on the research based on the framework of learning analytics remains open because different empirical results and studies differ largely, and it is important to know which indicator can expand the potential of learning analytics for meaningful analysis of pedagogical data.

Shum (2012) and Xing et al., (2015) discussed about the predictive ability of learning analytics and its impact at the different levels (micro, meso and macro) it operates. Shum (2012) further added that if learning analytics are being implemented systematically with the necessary academic models and pedagogical design principle, then it can have a promising impact on higher education. Relevant to this, Scheffel et al., (2014) pointed out how the quality indicator variables as shown in Figure 2, can be used as a standardized framework to evaluate learning analytics in educational practice.

![Figure 2: Quality Indicators for Learning Analytics by Scheffel et al., (2014)](image)

**Discussion and Conclusion**

In this paper we tried to look at the literature to inform on what has been done so far in terms of the potential of learning analytics to understand and model learner engagement in online courses and to subsequently look at how learner engagement can effectively be used as an indicator to predict performances. As can be seen from the papers reviewed, much of the engagement models of learning analytics rely on LMS data such as platform access logs, clicks and views, and also on completion of learning tasks such as participation in forums and completing online quizzes. However, there is a lack of conclusive evidence as researchers reported mixed findings with the pertinence of access logs to predict performances. On the other hand, participation in discussion forums and scores in class tests and assignments were found to be reliable predictors of final result in a number of studies. As some researchers have rightly stressed upon, while using learning analytics, one should not neglect the learning and pedagogical process which remains at the centre of any educational transaction. While analytics has mainly been used to track what a learner is doing and to relate it to a probability of success or failure, there is still room for research to determine what are the determinants of significant learning, and how to capture this information from the environment is important. For example, an important question that has to be addressed is whether more clicks can be equated to more learning, thus infer a higher probability of success? Or whether more logins or longer online sessions means that learning is actually happening? In an outcomes-based environment, the variables to be measured have to be defined in such a way that it is linked to the expected learning outcomes. This brings us to reflect deeper on what student engagement means in online learning.

Engagement can be in the form of access and clicks, or number of downloads, or the number of posts made by a student, but engagement can also be the time spent on a particular resource and taking down notes regularly and in an organised way either offline or online. Engagement can also mean the time spent by the student ‘outside’ the platform in hyperspace to further document himself or herself on the course components. Learning preferences of the learner can play an important role. For example, a socially-engaged learner might be expressing himself
more online while a reflective learner might prefer to spend more time to understand in more depth and participate less frequently online. As it has further been highlighted, not all online courses are designed in the same way. Some are centred around learning activities where frequent platform access is not necessary and a few sub-activities can be carried out either offline or on a computer. There is a need therefore to conceptualise student engagement in such courses, as platform logs and clicks alone would not suffice. An activity might also span over say 2 weeks where the learner has to read a few resources online, or offline, then watch one or two videos, then say use a software to develop a website and finally upload the work on the eLearning platform by the activity deadline. Self-reporting instruments such as the NSSE might contribute, but the reliability of self-reporting instruments has to be looked into. Therefore, there is a need to further reflect on how to define student engagement in such settings.

It is obvious that descriptive analytics allows the teacher to get first-hand information on students who are not regular on the eLearning platform and to design interventions to ensure that proper follow-up is done to ensure retention. Such functionalities have always existed in eLearning platforms like Moodle, and such information only works if the course designers have thought of a robust mechanism of student monitoring and support framework which is in any case an important quality assurance mechanism in distance online learning. Finally, the whole idea of learning analytics becomes more relevant only if the right data is available in big quantities as shown by the numerous studies. This is where the power of learning analytics lie, if it can help researchers to understand trends at macro-level, therefore helping in situations where there is a need for mass-customisation while on the other hand, use the predictive analytics through machine learning and other intelligent techniques, to single out cases where special attention is needed to achieve personalised learning.
References


[54] Zotou, M., Tambouris, E., Triantafyllou, E., Timcenko, O., Busk Kofoed, L., Stracke, C.M., Riviou, K., García Barriocanal, E., Utz, W., Martos, P. and Tarabanis, K., 2017. PBL3.0: Integrating Learning Analytics and Semantics in Problem-Based Learning. EC-TEL.