A systematic literature review in Learning Analytics

Andrés F. Pineda[1], Julian M. Cadavid[1]

[1]Department of Computer Science and the Decision– National University of Colombia
Cr 80 No 65-223 - Campus Robledo – Medellín – Antioquia – Colombia

`{afpinedac,jmoreno1}@unal.edu.co`

Abstract. *This paper presents a systematic literature review in Learning Analytics with the aim of providing a general overview for researchers about the current state of this area and in which educational contexts it has been applied. A five-step process was followed. First, a series of research questions were proposed in relation to the objectives, models, techniques and contexts in order to obtain an overview of the topic and its applications. Those questions allowed to direct the search to find relevant documents in this area, then a filtering process was carried out, followed by an evaluation process and finally, the analysis of the information to answer the questions proposed initially.*

1. Introduction

Learning Analytics is a fast growing area in the field of Technology Enhanced Learning (TEL). It has its roots in a variety of areas, mainly business intelligence, web analytics, data mining and recommendation systems (Ferguson, 2012). This is a method that can drastically improve the existing educational environment and increase the efficiency of management in educational institutions and in the teaching-learning process (Kim & Ahn, 2016).

One of the most accepted definition by the community is given in (Fernández-Gallego, Lama, Vidal, & Mucientes, 2013) where it is defined as "... *the discipline whose objective is the measurement, collection, analysis and reporting of data about the trainees and their contexts, in order to understand and optimize learning and the environment where it occurs*." (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-Garcíaa, 2014) describes it as "... *the analysis of data which allows professors, course designers and administrators of virtual learning environments to look for unobserved patterns whose objective is to improve learning outcomes*". (M. Chatti, Dyckhoff, Schroeder, & Thus, 2012) (describes it as "... *the use of intelligent data, data produced by the student and models of analysis to discover information, social connections and to predict and advise on learning*".

Learning Analytics is an area that has been the focus of research in different countries of the world. It currently has communities such as SOLAR (Society for Learning Analytics Research) in North America, LACE (Learning Analytics Community Exchange) in Europe, LALA (Learning Analytics Latin America) in Latin America. Likewise, it has specialized congresses such as LAK (Learning Analytics & Knowledge), SHEILA and specialized journals in this area such as JLA (Journal of Learning Analytics).

Different authors as (Verbert, duval, Klerkx, Govaerts, & Luis Santos, 2013), (Papamitsiou & Economides, 2014), (Avella, Kebritchi, Nunn, & Kanai, 2016), (Cechinel, Dos Santos, Carvalho, & Ochoa, 2017), (Bodily & Verbert, 2017), (Schwendimann et al., 2017) have presented systematic literature reviews in this area. These reviews are regularly focused or restricted to a particular location or Learning Analytics objectives. However, they are different to the approach proposed in this paper since the main goal of this review is to provide a general overview for future researches and the educational contexts where it has been applied.

2. Method

The systematic review of the literature was based on the procedure suggested by (Cooper, 1985). This procedure is based on a 5-step process: (a) formulate the problem, (b) collect data, (c) evaluate the appropriateness of the data, (d) analyze and interpret relevant data, and (e) organize and present the results.

2.1. Formulating the problem

One of the main objectives of this research is to have an overview of Learning Analytics and its applications. Therefore, the following conductive questions were proposed in order to guide the search for information:

• Q1: What are the main objectives in Learning Analytics?

• Q2: Which models / methods have been used as a reference to apply Learning Analytics?

• Q3: What techniques / algorithms have been used to meet these objectives?

• Q4: What have been the domains and target population in which Learning Analytics has been applied?

• Q5: Was it validated in a real environment, how and with what results?

2.2. Data collection

The bibliographic search was carried out through Science Direct, Scielo, Scopus, IEEE and Springer Journal databases. Likewise, the content platforms ACM Digital Library, Taylor & Francis Group, the Google Scholar search engine and the Web of Science metasearch engine were reviewed. The combination of key words used in the search were: "Model Learning Analytics", "Objectives Learning Analytics", "Techniques Learning Analytics", "Applications Learning Analytics", and "Survey Learning Analytics". Specific criteria for the selection of bibliographic material were established: the year of publication (between 2007 and 2018), and the type of document (article or book chapter). The location of the study was not taken in consideration as a selection criterion. The articles that met the established search criteria were downloaded and analyzed to extract information that allowed generating statistical reports.

2.3. Data evaluation

Based on the process described above, 189 items were found. Then a debugging process was done and some articles were discarded since they were not related to the approach

or did not respond to the initial questions. After this process, the total number of resulting articles was 129. The full list of articles can be found in http://tiny.cc/nc2oyy

2.4. Analysis and interpretation

The 129 papers have authors from 36 countries, as can be seen in Figure 1. Figure 2 shows the distribution of them by year of publication.
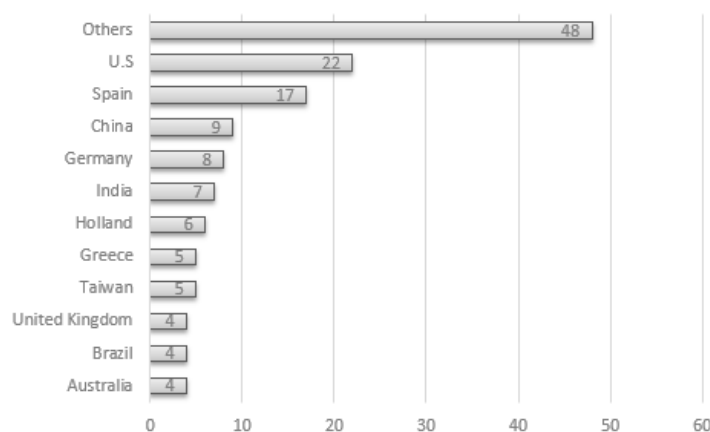

Figure 1. Publications by country

It is noteworthy a pattern of growth in the number of publications each year. Between 2005 and 2009, only one publication is submitted, whereas from then on, a considerable increase is seen.
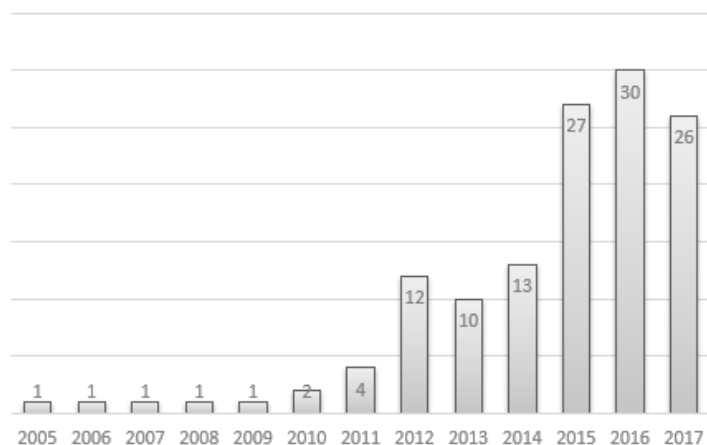

Figure 2. Number of publications per year

Finally, it is important to remark that 120 of these works corresponded to articles and 9 to book chapters. Being Computers & Education, Computers in Human Behavior and Educational Technology & Society the journals with the largest number of documents.

2.5. Results

After reviewing and analyzing the documents found, the initial questions are answered.

Q1: This question is related to why or for what purpose Learning Analytics is done. According to (Moissa, Gasparini, & Kemczinski, 2015), generally when Learning Analytics is done, it is done with the purpose of one of these 7 objectives:

- *Adaptation*: Refers to the process of dynamically restructuring the presentation of learning resources according to the needs of the student.

- *Evaluation and feedback*: Aims to evaluate learning outcomes by providing feedback.

- *Monitoring and analysis*: Its objective is to monitor the actions of the student (or any other interested party) and generate reports with the results of the observation.

- *Personalization and recommendation*: Allows students to create their own learning path by recommending resources or tutors based on their preferences or activities.

- *Prediction and intervention*: Consists in the creation of models to better understand and predict the performance of students based on their history, which indicates to students that they may need better attention and specific interventions.

- *Reflection*: Provides means to compare data between students, courses and professors to verify the effectiveness of learning practices.

- *Mentoring and tutoring*: Aims to help the student during a specific task or during the entire course.

Q2: One of the most widely accepted models is presented in (Moissa et al., 2015) (Atif, Richards, Bilgin, & Marrone, 2013). Which mention that the Learning Analytics process can be divided into 4 dimensions: What, Who, Why and How (Figure 3).
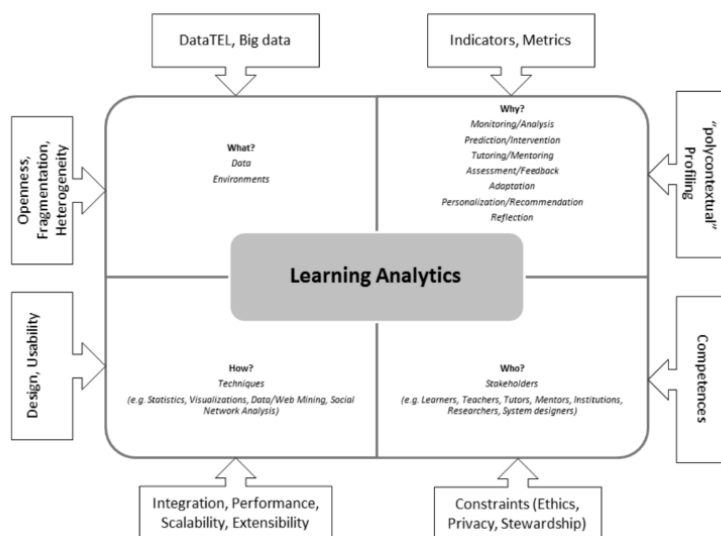
Figure 1. Learning Analytics model

(Source: (M. A. Chatti, Dyckhoff, Schroeder, & Thüs, 2012))

- *What*: Related to the data collected during the investigation. The source of the data can be virtual learning environments (VLE), surveys, social networks, electronic libraries or any other repository.

  Figure 4 shows that most of the research has been based on LMS data (39.2%). We found a considerable number of articles that were supported by Surveys (14%), Immersive Learning Environments (7.6%), Chats (7.0%), Online Discussion Forums (6.4%), Academic Information (5.3%). While Face-to-Face Interaction Supported by Technology (4.7%), Social Networks (3.5%), Video (2.9%), Haptic Means and Tangible Computing (2.3%), Intelligent Agents (2, 3%), Mobile Platforms (1.8%), Tutors (1.2%), Blogging (0.6%) were the least used during the investigations.
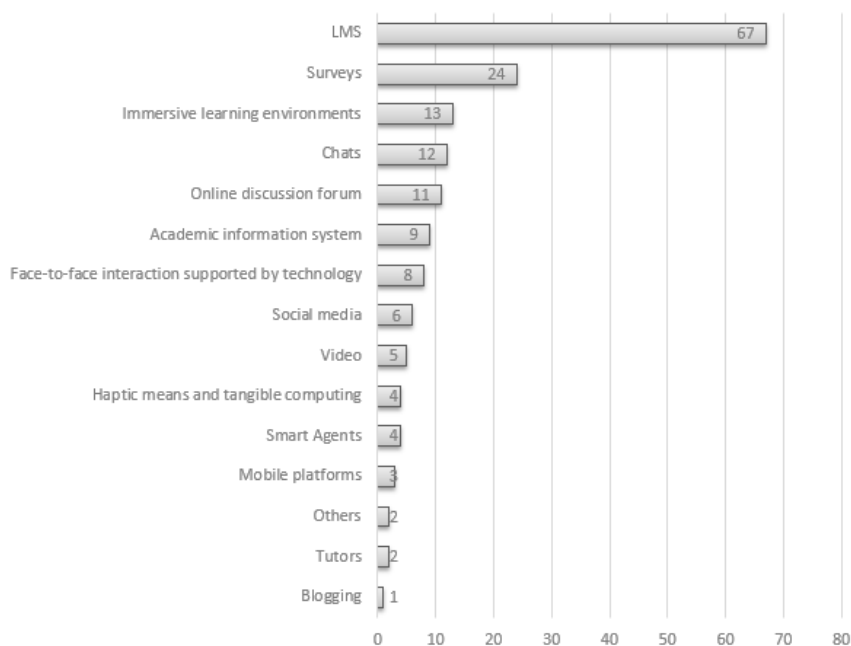
Figure 2. Distribution of works by types of data (*What* dimension)

- *Who*: It refers to the interested parties in the analysis. These can be professors, students, coordinators, researchers and institutions. In the present review, within this dimension a differentiation was made between the student population to which the analysis was directed, that is, university students and college students. In turn, a differentiation was made for specific groups within each of the two populations, that is, the number of elementary and secondary students (college), and undergraduate and postgraduate (university) students analyzed was determined.

  Of the 129 works, 34.1% of the publications are directed mainly to Teachers, followed by Students (23.3%), Institutions (9.2%), Researchers (22.5%), and Coordinators (10.8%). In relation to the student population, 80.7% of the work was directed to university students and 19.3% to college students. For the specific case of university students, it was found that 70.5% and 10.2% of publications were focused on the levels of undergraduate and postgraduate training, respectively. In the case of college students, 4.5% of the research was aimed at primary school students and 14.8% at high school students.

- *Why*: This dimension is related to the objectives of the analysis, divided into 7 categories: Adaptation, Evaluation and Feedback, Monitoring and Analysis, Personalization and Recommendation, Prediction and Intervention, Reflection, Mentoring and Mentoring.
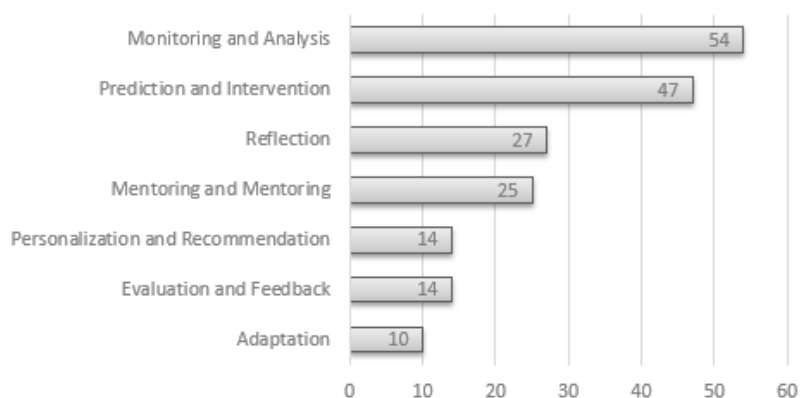
Figure 3. Distribution of works by objective (*Why* dimension)

Figure 5 shows the results corresponding to the dimension Why, most of the works (28.3%) are included in the Monitoring and Analysis category. Prediction and Intervention (24.6%), Reflection (14.1%) and Mentoring and Tutoring (13.1%), although to a lesser extent, they also turn out to be centers of interest in the developed works. On the other hand, Personalization and Recommendation (7.3%), Evaluation and Feedback (7.3%); and Adaptation (5.2%) are the least addressed categories.

- *How*: It is related to the techniques used during the investigation to achieve the objectives established in the dimension Why (Figure 6).

The articles were classified according to the four dimensions studied in this work and the results are shown below. It should be noted that some articles fit two or more categories of a certain dimension, so they were considered more than once.
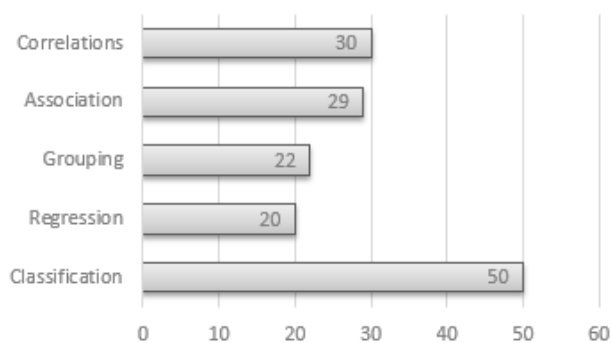

Figure 4. Analysis type

Q3: In Figure 7, the techniques used by researchers during their work to address the focuses of interest established in the Why dimension are shown. The results suggest that the most used technique is Descriptive Statistics (19.6%), followed by Decision Trees (11.9%).
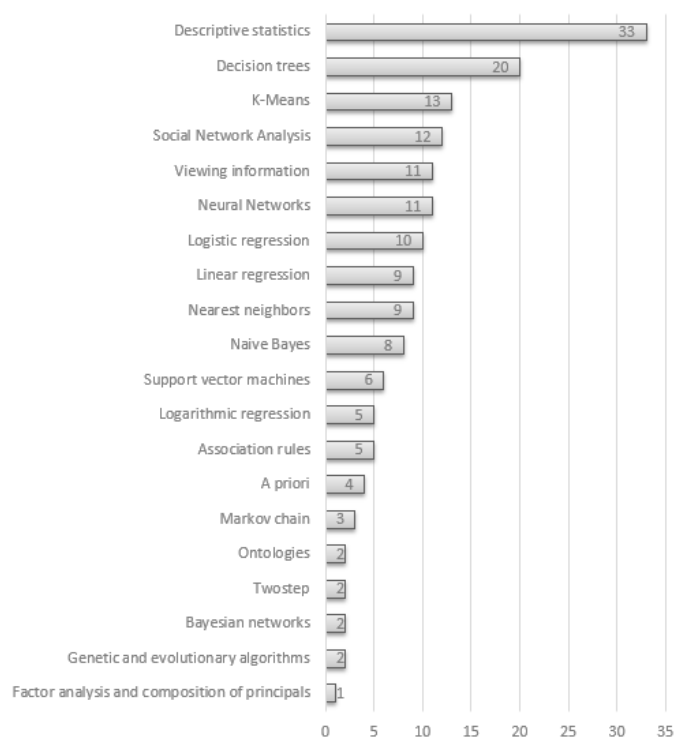
Figure 5. Distribution of works by technique used

Q4: Figure 8 shows the area of knowledge in which the research works were developed. For the area of engineering, for example, the vast majority of publications, 26, were focused on Programming Logic, 2 in Software Engineering, 1 in Networks and Telecommunications, 1 in Formal Languages, 1 in Computer Architecture, and 1 in Databases. On the other hand, in Mathematics and natural sciences, 14 were in Mathematics, 12 in Biology, 2 in Algebra, and 1 in Statistics.
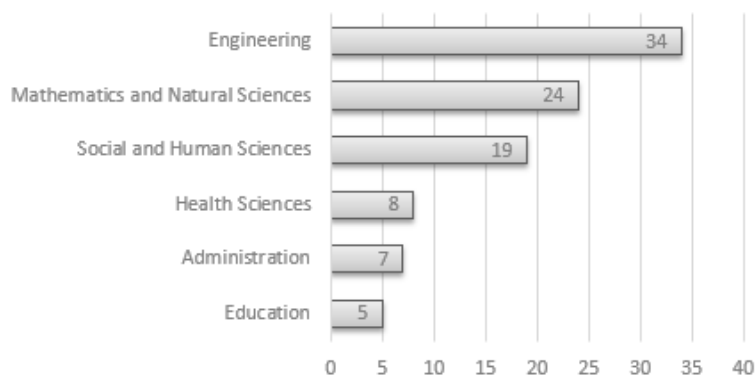

Figure 6. Distribution of works by context

Q5: This analysis allows us to evaluate how advances in research are reflected in educational practice, thus contributing to the improvement of education. It was found

that 66.7% of the works were validated in real environments, while 33.3% were not implemented, or were implemented, but not validated.

3. Conclusions

In this paper, a systematic review of the literature in the field of Learning Analytics was presented following a 5-step process in order to know a general overview in this area and its applications. In total, 129 documents were analyzed between articles and book chapters.

Based on the results obtained, a clear trend of growth can be seen in the number of publications that are made on this topic each year, mainly in the English language. Authors from the United States and Europe are the ones who, according to the results, have made the greatest number of contributions.

According to our analysis, the objectives of Learning Analytics can be divided into 7, with Monitoring and Analysis being the main one, followed by Prediction and intervention. To make this type of analysis most of the research makes use of tools such as Learning Management Systems, followed by surveys to obtain the required information.

The results also show that researchers generally validate their hypotheses through experimental courses and platforms that mainly help students and teachers to improve teaching-learning processes. Followed by a considerable number of works that presented some type of hypothesis or model that serves other researchers as a starting point for other investigations. And finally, they are interested in showing results in relation to results or policies of change in educational institutions.

Another interesting discovery that is made is that the majority of investigations have a focus on the university context, mainly in undergraduate, and in specific areas of programming and mathematics. This may be due to the facility of teachers to validate their hypotheses in courses in which they have greater knowledge.

In a future work, the possibility of making an exhaustive demographic and technical analysis of Learning Analytics is proposed. This is to analyze documents by language and review existing tools in this area.

References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-Garcíaa, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, *31*(1), 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Atif, A., Richards, D., Bilgin, A., & Marrone, M. (2013). Learning analytics in higher education: a summary of tools and approaches. *30th Australasian Society for Computers in Learning in Tertiary Education Conference (Ascilite 2013)*, 68–72.

Avella, J. T., Kebritchi, M., Nunn, S., & Kanai, T. (2016). Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learning Journal*, *20*(2), 13–29.

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research.

Cechinel, C., Dos Santos, H., Carvalho, J. B., & Ochoa, X. (2017). An Initial Review of Learning Analytics in Latin America. *XII Latin American Conference on Learning Technologies*.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A Reference Model for Learning Analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318–331. https://doi.org/DOI: 10.1504/IJTEL.2012.051815

Chatti, M., Dyckhoff, A., Schroeder, U., & Thus, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*, 318–331.

Cooper, H. (1985). A taxonomy of literature reviews. *American Educational Research Association*, *Chicago*, *I*.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5–6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Fernández-Gallego, B., Lama, M., Vidal, J. C., & Mucientes, M. (2013). Learning analytics framework for educational virtual worlds. *Procedia Computer Science*, *25*, 443–447. https://doi.org/10.1016/j.procs.2013.11.056

Kim, Y. H., & Ahn, J. H. (2016). A Study on the Application of Big Data to the Korean College Education System. *Procedia Computer Science*, *91*(Itqm), 855–861. https://doi.org/10.1016/j.procs.2016.07.096

Moissa, B., Gasparini, I., & Kemczinski, A. (2015). A systematic mapping on the learning analytics field and its analysis in the massive open online courses context. *International Journal of Distance Education Technologies (IJDET)*, *13*(3), 1–24.

Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, *17*(4), 49–64.

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., … Dillenbourg, P. (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522

Verbert, K., duval, erik, Klerkx, J., Govaerts, S., & Luis Santos, J. (2013). Learning Analytics Dashboard Applications. *American Behavioral Scientist*, *57*.