Chapter 1

LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT

Francisco José García-Peñalvo

Abstract: Learning analytics has become a reference area in the field of Learning Technologies as the mixture of different technical and methodological approaches in the capture, treatment and representation of educational data for later use in decision-making processes. With approximately ten years of development, it can be considered that learning analytics have abandoned their stage of dispersion and are heading towards a state of maturity that will position them as a fundamental piece in educational practice mediated by technology. However, it cannot be ignored that the power and goodness of these analytics must be channelled to improve learning itself and, therefore, the learning-teaching process, always acting from an ethical sense and preserving the privacy of the people who participate because it is straightforward to invade personal spaces in favour of the objectives sought. This chapter presents, from a conceptual perspective, the reference models that support the creation of educational strategies based on learning analytics that integrate the most current trends in the field, defined from a critical perspective that balances the undoubted benefits with the potential risks.

Key words: Learning analytics, Educational improvement, Reference models, Trends, Risks
1. INTRODUCTION

We live in a data-driven society. Data is the new asset for many social stakeholders, such as politicians, decision-makers, journalists, and so forth. This fact is related to the automatic treatment and analysis of the acquired data to enhance the value of the critical procedures in the organisations.

Education, or academia in general, is no exception to this dependence on data, with the widespread excuse of improving learning outcomes and experience within an increasing proliferation of online offered courses, with an evident influence of the Massive Open Online Courses (MOOC) (García-Peñalvo, Fidalgo-Blanco, & Sein-Echaluce, 2018) or the blended modes of learning (Graham, 2006). Since 2010 approximately, learning analytics (LA) has been considering as an emerging research line in the Learning Technology area (Johnson, Smith, Willis, Levine, & Haywood, 2011; Siemens, 2013).

The fundamental core of LA is based on methods that harness educational data to support the learning process, taking into account that many types of educational data exist, with different characteristics such as distribution, scope, size, and privacy (Almosallam & Ouertani, 2014). Moreover, LA is related to other fields such as Educational Data Mining (EDM) (Romero & Ventura, 2007, 2010), which is oriented to gather the data, or Academic Analytics (AA) (Campbell, DeBlois, & Oblinger, 2007; Goldstein & Katz, 2005), which is focused on business intelligence practices for analysing academic data with administrative goals to make institutional decisions.

LA uses different techniques to make the analytics processes, such as Artificial Intelligence (AI) based methods (García-Peñalvo, Cruz-Benito, et al., 2018; Lu & Hsiao, 2019); visual analytics approaches (Gómez-Aguilar, Hernández-García, García-Peñalvo, & Thrón, 2015; Gómez-Aguilar, Thrón, & García-Peñalvo, 2009; Villamañé, Álvarez-Larrañaga, Caballero, & Hernández-Rivas, 2018), dashboards (Vázquez-Ingelmo, García-Peñalvo, & Thrón, 2019; Vázquez-Ingelmo, García-Peñalvo, Thrón, & Conde, 2019), data visualisation techniques (Leony et al., 2012; Verbert, Duval, Klerks, Govaerts, & Santos, 2013), multimodal analytics (Worsley, 2018) or mixed quantitative and qualitative methods (Munce & Archibald, 2016; Palomo Duarte et al., 2018), among others.

The objectives of the LA approaches are as varied as the needs that stakeholders in the educational process have to make decisions (Conde-González & Hernández-García, 2015) based on the academic data collected, such as, monitoring and analysis (Conde-González, Hernández-García, García-Peñalvo, & Sein-Echaluce, 2015), prediction and intervention (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Fancsali et al., 2018; Liz-Domínguez, Caeiro-Rodríguez, Llamas-Nistal, & Mikic-Fonte, 2019), assessment and feedback
1. LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT

(Amo-Filvà, Alier Forment, García-Peñalvo, Fonseca-Escudero, & Casañ, 2019; Fidalgo-Blanco, Sein-Echaluce, García-Peñalvo, & Conde-González, 2015), intelligent tutoring (Doroudi & Brunskill, 2019), adaptation, personalization and recommendation (Jiang, Pardos, & Wei, 2019; Mangaroska, Vesin, & Giannakos, 2019), retention (Feild, Lewkow, Burns, & Gebhardt, 2018), massive online training (Andres et al., 2018; Cobos & Macías Palla, 2017; Lerís, Sein-Echaluce, Hernández, & Fidalgo-Blanco, 2016), and so on.

However, faced with the many potential advantages associated with automatic analysis of educational data, important questions arise, primarily related to the ethical and privacy (Alier Forment, Amo Filvà, García-Peñalvo, Fonseca Escudero, & Casañ, 2018) aspects of the teaching/learning process.

Besides, from a deeper perspective, the past and present development of LA is generally more concerned with trying to correlate the online activity count with academic performance than with demonstrating a long-term impact on online students and teaching practice (Gašević, Dawson, & Siemens, 2015).

From the consumer perspective of the outcomes, results, and reports of the LA tools, we find two different and related problems, users need a deeper visual literacy (Felten, 2008) to understand the offered data visualisations and the designs of the dashboard are usually far less to be understandable (Tanes, Arnold, King, & Remnet, 2011).

Several reference models and frameworks try to relate all the characteristics, components, methods, benefits, and issues that comprise the field of LA.

This chapter tries to give a conceptual view of the current state of the LA practice to discuss if this means an actual improvement in the technology-enhanced learning. To do that the section 2 includes the most usual definitions of LA; section 3 present the most important reference models for LA; and section 4 discusses the current state of the LA development.

2. LEARNING ANALYTICS DEFINITIONS

In a conceptual paper about LA, the most significant definitions of the discipline should be presented.

Most of LA literature has adopted the definition offered in the 1st International Conference on Learning Analytics, LAK’11, held in Banff, Alberta (Canada) on February 27-March 1, 2011 (https://tekri.athabascau.ca/analytics/) and adopted by the Society for Learning Analytics Research (SoLAR):

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 environments in which it occurs.
George Siemens (2010) makes a highly referenced definition of LA in his blog:

The use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.

The Horizon Report for Higher Education has included LA in different editions. In the 2011 Report the following reference is provided (Johnson et al., 2011):

Learning Analytics refers to the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues.

Elias (2011) defines LA as:

An emerging field in which sophisticated analytic tools are used to improve learning and education.

Educause proposes the following LA definition (van Barneveld, Arnold, & Campbell, 2012) (as an adaptation of the Bach’s definition (2010)):

The use of analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals.

The proposed definitions present differences, but also share the mission to use educational data in actions to improve learning, with a particular emphasis on the student or learner, but including relationships with some institutional level.

As it was mentioned in the introduction, the emergence the LA as independent field means we distinguish three main analytics areas with an existing overlapping among them: EDM focused on the technical challenge: How can we extract value from these big sets of learning-related data?; LA focused on the educational challenge: How can we optimise opportunities for online learning?; and AA focused on the political/economic challenge: How can we substantially improve learning opportunities and educational results at national or international levels? (Ferguson, 2012).

Long and Siemens (2011) establish a clear difference between LA and AA, where LA benefits learners and faculty, and it is focused on course and department levels, whereas AA benefits decision-makers at institutional, regional, and national levels.

3. LEARNING ANALYTICS REFERENCE MODELS

The analytical process is an iterative cycle that usually comprises five main steps: 1) data collection; 2) data pre-processing; 3) analytics; 4) post-processing; and 5) decision-making. These steps are thought to be supported by an automatic analytics process; however, decision-making requires human participation based on the visualisation of the processed data. When the user may interact with the data, a visual analytics process appears, which is a complementary process to the automatic one.
LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT

Figure 1 summarises the LA process combining the basic steps (Chatti, Dyckhoff, Schroeder, & Thüs, 2012) with the visual analytics capabilities (Keim et al., 2008).

Taking into account this process model, we are going to explore the most important reference models for LA.

3.1 Chatti et al. reference model for Learning Analytics

Chatti et al. (2012) define a four-dimensions reference model (see Figure 2). The identified dimensions are:

- What? What kind of data does the system gather, manage, and use for the analysis? In this dimension, where the educational data comes from is one of the most important questions to answer due to the different sources we manage in an educational context. In the current configuration, monolithic information systems have passed through, the typical Learning Management System (LMS) (García-Peñalvo & Seoane-Pardo, 2015; Gros & García-Peñalvo, 2016) is still alive, but as another component of a complex technological ecosystem (García-Holgado & García-Peñalvo, 2019; García-Peñalvo, 2018).

- Who? Who is targeted by the analysis? The LA is devoted to being applied toward different stakeholders: students, teachers, tutors, institutions, researchers, and system designers with several perspectives, goals, and expectations.
• **Why?** Why does the system analyse the collected data? If we have different stakeholders, there exist many objectives according to the particular perspective of every involved stakeholder.

• **How?** How does the system perform the analysis of the collected data? LA applies different techniques to detect interesting patterns within the educational data sets.

![Learning Analytics diagram](image)

**Figure 2. Chatti’s Learning Analytics reference model, adapted from (Chatti et al., 2012)**

### 3.2 Greller and Drachsler framework for Learning Analytics

Greller and Drachsler (2012) define a six-dimensions framework for learning analytics that are explained in Table 1.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stakeholders</strong></td>
<td>It includes data clients (beneficiaries of the LA process who are entitled and meant to act upon the outcome, e.g., teachers) as well as data subjects (they are suppliers of data, normally through their browsing and interaction behaviour, e.g., learners).</td>
</tr>
<tr>
<td><strong>Objective</strong></td>
<td>The main opportunities for LA are to unveil and contextualise so far hidden information out of the educational data and prepare it for the</td>
</tr>
</tbody>
</table>
### 1. LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Educational data</strong></td>
<td>LA takes advantage of available educational datasets from the learning technological ecosystem. Educational data might be classified as open and protected datasets.</td>
</tr>
<tr>
<td><strong>Instruments</strong></td>
<td>Different technologies, techniques, and tools can be applied in the development of educational services and applications that support the objectives of educational stakeholders.</td>
</tr>
<tr>
<td><strong>External constraints</strong></td>
<td>Many different kinds of constraints can limit the beneficial application of LA processes. The most significant ones are ethical, legal, and social constraints.</td>
</tr>
<tr>
<td><strong>Internal limitations</strong></td>
<td>A number of human factors that enable or may pose obstacles and barriers, prominent among these are competences and acceptance.</td>
</tr>
</tbody>
</table>

The dimensions of this framework are shown in [Figure 3](#). Each dimension is divided in a set of instantiations. The list of instances in the figure does not pretend to be an exhaustive enumeration and can be extended. Greller and Drachsler classify these dimensions as critical because all of them are required to have at least one instance in a fully LA design.

![Figure 3. Greller and Drachsler framework for Learning Analytics, adapted from (Greller & Drachsler, 2012)](#)
3.3 VeLA model

Gómez-Aguilar et al. (2014) define the Visual eLearning Analytics (VeLA) model. The roots of this model are in the belief that LA, AA, and EDM share the common goals of improving and understanding better the learning process, requiring a huge number of observations to do that.

Figure 4 presents the theoretical model of VeLA, showing how LA and AA have a common set of objectives and technological instruments, but the VeLA area complements and enriches both LA and AA processes. The bases of this model are the Chatti et al. (2012) reference model, the Clow (2012) LA loop theory, the Greller and Drachsler (2012) critical dimensions, and the Keim data visual exploration process (Keim, Kohlhammer, Ellis, & Mansmann, 2010; Keim & Zhang, 2011).

Shneiderman (1996) proposes an information-seeking paradigm that was extended by Keim et al. (2008) to give more capacity for analysis in the cycle before and after of generating interactive visuals representations. With these, the user might explore the data to extract abstract models from of data sets that are too large or too complex to be analysed in a direct way, in such a way that the user’s questions are answered or formulated during the exploration cycle. Also, the paradigm of Keim, applied to educational analytics, can be extend to a final step, the intervention, by providing to the analytical cycle a feedback, leaving the following form:

Analyse first; Show the important; Zoom, filter and analyse further; Details on demand; Intervention.

This cycle is represented in the central part of Figure 5. It is shown a general and abstract view of the stages (represented by circles) and the transitions (represented by arrows) in the VeLA process. Inside of each stage, both LA and AA are represented. Also, different stakeholders’ profiles are supported, everyone with a different goals and interest in the provided information.

The VeLA process presents visual analytics techniques to support temporariness, content semantic analysis, social network analysis, and statistics.
1. LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT

Figure 4. VeLA theoretical model (Gómez-Aguilar et al., 2014)

Figure 5. eLearning visual analytics process (Gómez-Aguilar et al., 2014)
4. DISCUSSION

Learning analytics, or analytics in general in the educational context, are called to play a crucial role in the teaching/learning process because they allow having evidence to make decisions aimed at improving learning.

In the last decade, many of the research efforts in the field of learning technologies have focused on the development of tools, methods and techniques to support the analysis of educational data.

During these years, many works have been carried out that have been constructing the reference body of analytics in educational processes, logically from some stages of greater dispersion to others of greater maturity, following the principles established in the definitions and the frameworks or reference models.

The incidence of analytics in the teaching/learning process can be oriented, in a general way, to two pivotal moments.

On the one hand, actions that are carried out when the educational process has finished or is going to be finished, for example, to provide automatic or semi-automatic support to a summative evaluation of the students or to carry out global analyses oriented to decision making for future formative initiatives.

On the other hand, actions that take place while the teaching/learning process is being developed, and can modify it, such as, for example, feedback and recommendations to students, predictive warnings to the teacher about the progress of the process, or support for the personalisation of learning (Lerís & Sein-Echaluce, 2011) or adaptive learning (Berlanga & García-Peñalvo, 2008; Burgos, Tattersall, & Koper, 2007; Lerís, Sein-Echaluce, Hernández, & Bueno, 2017).

Regardless of the interest in learning analytics, MOOCs and the development of artificial intelligence in education have focused on the analysis of educational data, for different purposes, but with a particular emphasis on personalisation/adaptation of learning.

The xMOOC model has several limitations (Fidalgo-Blanco, Sein-Echaluce, & García-Peñalvo, 2016) that could be solved or mitigated introducing analytics to make decisions or personalise the learning paths (Yu, Miao, Leung, & White, 2017), splitting the total number of students in smaller groups in which to apply other educational approaches (Sein-Echaluce, Fidalgo-Blanco, & García-Peñalvo, 2017; Sein-Echaluce, Fidalgo-Blanco, García-Peñalvo, & Conde-González, 2016).

Artificial intelligence methods might be used for different support tasks in the overall cycle of managing educative data, from mining the sources to the automatic treatments and decision-making processes. However, one of the most controversial issues is the idea of substituting the teachers by automatic bots and intelligent tutors (Farhan et al., 2012; Frank, Roehrig, & Pring, 2017).
Obviously, the big problem is the tendency to generalization. A scenario oriented to a training in which the student-student or student-teacher interaction is not necessary to achieve the learning objectives can be an ideal field for the use of intelligent agents that accompany the student in his learning process. To derive from these scenarios that any educational activity can dispense with the teaching staff would be to fall into demagoguery.

If we use the following equation to describe a learning strategy:

\[ \text{LA + AI + Personalisation} - \text{Human effort} \]

We have a training activity in which the information component would be higher than the educational intention, with few or none collaboration and interaction. In this scenario, the learning goals and outcomes might be reached, and it might be oriented to unsupervised training actions.

On the other hand, if we use this equation to describe another learning strategy:

\[ \text{LA + AI + Personalisation + Human effort} \]

We have an educational activity in which interaction and collaboration might be high, reducing the teacher effort by introducing automatic and intelligent tools and agents in the learning ecosystem. This scenario combines the best of having people leading the teaching and learning process supported by robust and intelligent technology.

These two scenarios are represented in Figure 6.
LA + AI + Personalisation – Human effort

- Training
- Information
- Education
- Collaboration
- Interaction

LA + AI + Personalisation + Human effort

- Training
- Information
- Education
- Collaboration
- Interaction

Learning

Figure 6. Learning strategies combining LA, IA, personalisation and the human effort

LA-based technologies, tools, methods, and strategies will be an actual breakthrough in education when their maturity reflects that they are included transparently in the technological learning ecosystem and the learning design workflow with a central goal, improve the learning and the teaching and learning process, with an entirely ethical behaviour that preserves the privacy of the involved stakeholders.

REFERENCES


1. LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT


1. LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT

assessment. Computers in Human Behavior, 47, 149-156. doi:10.1016/j.chb.2014.11.050


blended learning: Global perspectives, local designs (pp. 3-21). San Francisco, USA: JosseyBass/Pfeiffer.


1. LEARNING ANALYTICS AS A BREAKTHROUGH IN EDUCATIONAL IMPROVEMENT

ACKNOWLEDGEMENTS

This work was supported in part by the Spanish Ministry of Science, Innovation, and Universities throughout the DEFINES Project under Grant TIN2016-80172-R.

AUTHOR INFORMATION

García-Peñalvo, Francisco José

GRIAL Research Group
Computer Science Department
Research Institute for Educational Sciences
University of Salamanca

Facultad de Ciencias. Plaza de los Caídos, S/N
37008 Salamanca, Spain

+34 923294400 Ext. 6095

fgarcia@usal.es

https://grial.usal.es/fgarcia

Full Professor Francisco José García-Peñalvo received his bachelor’s degree in computing from the University of Valladolid (Spain), and his PhD. degree from the University of Salamanca, where he is currently the Head of the Research Group in Interaction and e-Learning (GRIAL). His main research interests focus on eLearning, computers and education and digital ecosystems. He is the Editor in Chief of the Education in the Knowledge Society journal and the Journal of Information Technology Research. He coordinates the Doctoral Program in Education in the Knowledge Society.