From Big Data to Learning Analytics for a personalized learning experience

Anna Dipace a, F. Feldia Loperfido b, Alessia Scarinci c1

aUniversità di Modena Reggio Emilia, Italy, anna.dipace@unimore.it, ORCID 0000-0001-9826-073X
bUniversità di Foggia, Italy, feldialop@gmail.com, ORCID 0000-0003-2392-0021
cUniversità degli Studi di Bari, Italy, alessia.scarinci@uniba.it, ORCID 0000-0002-3174-7137

Abstract

This article describes Learning Analytics (LA) as a predictive and formative approach that enables the planning of educational scenarios in line with students’ needs and languages in order to set a priori and in progress systems of control and inspection of the following: consistency, relevance and effectiveness of training objectives, curriculum paths, students’ needs and learning outcomes. Thanks to LA, it is possible to understand how students learn. Training courses are designed to include the definition of those learning outcomes that respond effectively to students’ needs in terms of contents, methodologies, tools and teaching resources. The present article aims to describe and discuss, after reviewing the relevant literature, in what way LA represents a valid support not only in designing student-centred training courses, which assess outcomes, but also in carrying out a formative assessment considering the learning experience as a whole. The analysis of some case studies was a good opportunity to reflect and define the bridge existing between the use of LA for assessment purposes and personalized learning paths.

Keywords: Learning Analytics; learning outcomes; personalization; predictive analytics; formative assessment.

1. Higher education and digital culture: a matter of data

Digitizing formal, non-formal and informal cultural and educational contents, as well as innovating educational offerings, has enriched universities with fully online, blended and Mooc courses, and they have also seriously undermined traditional pedagogy, training models, the identity and the mission of universities in the third millennium. Innovation has involved not only the methods and tools, but also the different figures involved in such a profound change of higher education in form and structure, above all students and teachers.

Students’ approach to culture, knowledge and college education has changed; expectations and assessment of the quality of teaching they are directly involved in have changed, as well. The professional identity of the teacher is at the center of important turning points, which see a paradigmatic shift from an essentially transmissive, linear, one-to-many approach, to a socio-constructionist one where the student is at the center of the educational project. As Antonella Lotti (2018, p.20) supports, the teacher is asked for a “planning skill for educational intervention that takes into consideration the level of preparation of the students and which knows how to encourage.” Teachers hold a vital role because they need to be able to plan their teaching based on an analysis of their students’ needs, in line with the pre-set learning objectives.

The shift towards learning environments that are accompanied by the word “technological” causes new problems to emerge and makes us reflect on the traditional distinction between formal, informal and non-formal educational places, in order to grasp the integration processes and to conceive subject and environment as two interacting parts of a whole (Paparella in Limone, 2012a, p.10). The development and the rapid diffusion of virtual learning environments (or virtual spaces where interaction takes place and where knowledge is built and co-built), where learning is designed and

1 This article has been developed jointly by the authors. Anna Dipace wrote the sections: 1 - Higher education and digital culture: a matter of data; 2. Use of predictive learning data and models. F. Feldia Loperfido wrote sections: 3. Learning Analytics as a formative assessment tool for personalized learning; 4 Description of some case studies. Alessia Scarinci wrote the section: 5. Discussion and conclusion.
shared among inter-connected global users has obliged society to face the challenge of the well-known big data. They are “sets of data whose size goes beyond the capacity of typical database software tools and are used for acquiring, storing, managing and analyzing” (Manyika et al., 2011, p.1 in Ferguson, 2014). Institutions operating in higher education and vocational training have found themselves managing a mass of data which grows progressively/exponentially as a result of user interactions, an increase in the input of personal data in the different systems and the integration of other multiple information.

This new field of investigation linked to the analysis of data has encouraged the emergence of particularly interesting fields for research and educational planning. Indeed, scientific research activities have started to spread in the field of pedagogical analysis techniques such as Educational Data Mining (EDM) (at a technical level), Learning Analytics (LA) (in education), Academic Analytics (AA) (in politics and economics) (Ferguson, 2014). What could the connection be between these three levels of analysis to support educational intervention? How does the role of the student change in the era of Big Data? What educational choices should be made to ensure that students do not turn into a mere accumulation of data, but that they find further support in their development process thanks to an ethical and smart use of that same data?

A first general answer to these questions comes from the definition of Learning Analytics, or from a statement of the objectives that the LA techniques should help attain. The LA level is a dimension capable of giving a psycho-pedagogical meaning even at the level of EDM and AA. Research in Learning Analytics, in fact, aims to build meaningful links with learning sciences, distinguishing them from analysis techniques and data mining fields, Academic Analytics and EDM (Ferguson, 2014).

Generally speaking, Learning Analytics is used to measure, collect, analyse and communicate data related to students and the contexts in which they learn (Loperfido, Dipace, Scarinci, 2018). The main aim of LA is to allow schools and training designers to:

- Understand, differentiate and optimize learning and the environments in which it occurs;
- Assess the learning experience and the outcomes both as ongoing and a posteriori;
- Customize educational opportunities at the various needs levels and the abilities of each student and in real time (NMC, 2012, p.22).

More specifically, the approach of LA is encapsulated in a series of possible tools suggested by Educational Data Mining, which, in our opinion, may make sense if they are filtered through the psychoeducational concepts of formative assessment (Cauley & McMillan, 2010; Kingston & Nash, 2011) and personalization (Cordova & Lepper, 1996).

2. Use of predictive learning data and models

The LA community came to be at the first International Conference on Learning Analytics, held in Banff in 2011 and was followed by the foundation of SoLAR (Society for Learning Analytics Research) in that same year. Early research in LA meant beginning didactic planning within virtual learning environments, invigorated by the promise of customizing training offerings in an even more focused way based on students’ interests, needs and performances. A first definition of LA goes back to Siemens, who defined it as “the use of intelligent data, of data produced by students, as well as analysis models to discover information and social connections, and to predict and give advice about learning” (Siemens, 2010, available online on http://www.elearn-space.org/).

This definition was, then, developed by subsequent discussion between international researchers and the SoLAR foundation which, exactly a year later, tried to reformulate it in order to understand the different perspectives that had emerged from the scientific debate. According to SoLAR, LA refers to “the measurement, collection, analysis and presentation of data to students and their contexts, for the purposes of understanding and optimizing learning and the environments in which this occurs” (LAK11, 2011). In this sense, it is possible to consider the different nuances characterizing LA, which make it useful for effective interventions. First of all, LA techniques have numerous advantages because, like business intelligence, they can make predictions of students’ behaviour in a specific virtual learning environment.

In business intelligence (Quagini, 2015), for example, forward-looking investigations are establishing themselves in order to anticipate events and, thus, obtain benefits in economic terms. Specifically, these are Data mining processes (Gorunescu, 2011) that allow researchers to go deep into the data and extract information, patterns and relationships that would not be immediately extracted, without using these specific detection techniques. The same analytical processes have also become of interest in educational contexts developing Educational Data Mining (EDM) techniques and Learning Analytics (LA), as well as their respective communities of practice (Papamitsou & Economides, 2014). Together, EDM and LA represent an “ecosystem of methods and techniques” (Papamitsou & Economides, 2014, p.49) through which it is possible to collect and process data that is not limited to the mere reading and collation, but which also facilitates in-depth analysis leading to the prediction of students’ behaviours and, therefore, contributes to the improvement of teaching processes and learning outcomes.

The focus of LA is the adoption of forward-looking models in education systems through the description of data and results using specific techniques, such as: statistics, SNA projection, sentiment analysis, influence analytics, discourse analysis, concept analysis, and sense-making models. Academic Analytics, as mentioned above, represents the adoption
of business intelligence in education and focuses on analytics at the international, regional and institutional levels (Long & Siemens, 2014). As Campbell, DeBlois, Oblinger (2007, p. 64) stated, “data analysis techniques combine large sets of data, statistical techniques and predictive modeling. They might be thought of as the practice of extracting institutional data to generate actionable intelligence,” that is dynamic, viable and adaptive intelligence. It goes without saying that the boundary between the use of data for business reasons and for encouraging more effective learning becomes particularly blurred. Therefore, in Learning Analytics, personalized objectives play a more important role than marketing, and tailor-made learning is more important than the commercialization of knowledge. They marry scientific fields of data science to pedagogical sciences with the aim of designing new models of systematic collection and progressive modeling of data towards Actionable Data with the general objective of solving specific educational problems and defining forward-looking analysis techniques (de Waal, 2017). These derive from the possibility of using Data Mining techniques to synthesize a vast amount of data into powerful decision making skills (Baker, 2007).

Learning Analytics, as Siemens & Becker (2012) argued, provide new techniques for reading data, bringing the focus of educational research closer to data-driven decision-making science, integrating technical and socio-pedagogical dimensions. In this sense, LA enables the analysis of educational processes in terms of assessment and quality of interactions. In doing so, pedagogical research is not limited to the analysis of learning outcomes but it uses data that allows ongoing monitoring of the training processes using “current and contextual” data (de Waal, 2017, p. 32).

As Ifenthaler (2015) affirms, LA provides the pedagogical and technological background to provide real-time interventions at any time during the learning process. This personalized, dynamic and timely feedback represents an important support for the students’ self-regulated learning process and influences their motivation and their subsequent levels of achievement. In this perspective, a “holistic LA framework” (Ifenthaler, 2015) is expected to provide evidence that can be used to increase the overall quality of learning environments and facilitate the reform of learning and teaching in the 21st century.

3. Learning Analytics as a formative assessment tool for personalized learning

The concept of personalized learning was born thanks to contributions by authors such as Bruner, Maria Montessori, Dewey and Clapèrède, who can be considered the founders of today’s psycho-pedagogy. Later on, this concept regains some substance and applicability thanks to the development of digital contexts that make constant monitoring of learning processes possible, and increases the feasibility of remodeling them in progress. In particular, learning can be defined as personalized when it ensures the following: objectives that differ for every student, active participation and self-direction of the students in building their own path, developing students’ different dimensions (cognitive, emotional, relational, etc.) and strengthening of previous knowledge and skills (Cordova & Lepper, 1996).

Therefore, if we wish to list some of the advantages of the introduction of such pedagogical and technological innovations in education, the following must be included (Fulantelli & Taivi, 2015, p.130):
- Assessment of individual student’s learning disabilities;
- Overall assessment of the course;
- Early detection of predictors of dropout risk in education.

In the 2011 Horizon report by New Media Consortium Horizon, LA is seen as “interpreting a wide range of data produced by students and collected on their behalf in order to assess academic progress, predict future performance and identify potential problems” (Johnson et al., 2011, p.28). LA is a powerful tool in formative assessment, as originally described by Michael Scriven (1967), according to whom LA is a kind of assessment action that has effects on the pedagogical process. Certainly, it is applied during the curriculum in order to improve its effectiveness. It acts directly on learning objectives and results, and provides feedback to encourage student progress. In fact, feedback can activate self-evaluation processes both between teachers and students, and, if necessary, it can spur a change in teaching or learning activities in which these actors are respectively involved (Limone, 2012b). The concept of formative assessment is used when analytical evidence is used to facilitate the adaptation of teaching to real learning needs (Black & Wiliam, 1998).

Feedback is a key element in teaching and learning. Valerie Shute (2008), in her research on formative feedback, identified the characteristics of effective formative feedback: for example, feedback should be non-evaluative, supportive, timely, targeted, multidimensional and credible. The availability of continuous and recursive feedback can influence the motivation to learn. In this sense, Learning Analytics provides the opportunity, which never existed before, to find out whether the integral aspects of the curriculum are working as planned, and to effect continuous monitor of the progress of the course in the light of its pre-set objectives (Sclater, 2017, p.61).

Early warning systems fall into place when measuring and assessing the learning objectives and the use of ongoing monitoring systems for the consistency, relevance and effectiveness of curricular pathways and learning outcomes. They are also known as “student success” or “early warning systems.” They are kinds of software that can collect data referring to student engagement, a variable that indicates students’ involvement in the learning process (Sclater, 2017).
Therefore, it is a tool for evaluating educational processes within educational institutions. This approach to assessing the quality of processes and training systems uses historical and educational data to identify students in real time who might be at risk of academic failure. The use of such systems, relying on “actionable intelligence” can favour a timely change in students’ behaviour (Lonn et al., 2015).

4. Description of some case studies

A well-known example is Course Signals used at Purdue University in Indiana to prevent drop-out (Arnold & Pistilli, 2012). The system adopted consists of a traffic-light signal used for all students to indicate their possible risk of failure. This tool represents a device that acts as an ongoing assessment tool for students, but it also assesses the quality of the processes for the institution.

The software product developed at the University, Course Signals is designed to increase student success by using analytics to alert faculties, students, and staff to potential problems.

In particular, at the student level, this LA system gives them feedback on the progress of their learning process. At the same time, students do not run the risk of receiving a negative evaluation when it is too late, and accordingly they have enough time to ask for help. In this way, dispersion can be reduced and corrective actions can be promoted through scaffolding strategies and formative feedback that leads students to improve their learning and their final grade.

At the institutional level, the goal is to improve overall retention and the academic success rate and, consequently, the number of students who graduate (Sclater, 2017). This device represents a traffic-light signal, which depending on the light (whether red, yellow or green), indicates the level of risk run by each student is at a certain point in his or her course of study. The predictive algorithm takes into account four components (Sclater, 2017, p.38):

- Performance: based on the grades obtained during the course up to a certain point;
- Effort: the level of interaction with the LMS environment compared to other students;
- Academic background: including the students’ average grades from high school and the standardized grades;
- Characteristics of the student (i.e. age).

Thanks to a formula that considers a variety of predictors and current behaviors (e.g. previous grade point average, attendance, running scores), Course Signals can help spot potential academic problems before traditional methods might. This formula allows the institution to verify the progress of the students in a specific course through a green-yellow-red scheme that clearly indicates whether the students risk the dreaded DWIF (dropping out, withdrawing, getting an incomplete, or failing).

The main goal of Course Signals, in a nutshell, is to help students become academically integrated with the institution, with the use of learning analytics enabling real-time data on student performance and interaction with the LMS to be combined with demographic information and data on past academic history (Arnold, Pistilli, 2012).

Case studies similar to that at Purdue University have been compiled, for example, at Marist College, Nottingham Trent University and the University of New England in Australia, just to mention a few examples (Scatler, 2017). At the same time, several tools have been developed and included in teachers’ and students’ dashboards to monitor the progress in terms of participation, performance, interaction, etc. These tools support both summative (i.e. ex post on learning outcomes) and formative (i.e. in progress on learning) assessment, and several studies have shown their effectiveness in terms of increasing motivation and involvement of students (Park & Jo, 2015).

More specifically, the Open Universities of Australia (OUA), which is a consortium of seven Australian Universities providing distance education, have proposed the use of the Personalized Adaptive Study Success (PASS) (Ifenthaler, 2014). This is a tool to help students recognize learning paths appropriate to them. In a traditional learning experience, the student participates in several and sequential modules. Therefore, low performances on one module impact on the assessment of the subsequent modules. However, through PASS, a student struggling with a particular topic can study extra material on the subject to reinforce that knowledge before re-assessing the original module. PASS reads and elaborates data from different sources, such as the curriculum profiles of each module, the learning management system of the course and the OUA’s customer relationship management system. That is, it looks at the data produced by the student profile (his socio-demographic characteristics, prior learning experiences and so on), the learning profile (assessments, the use of online forums and contents, etc.) and the curriculum profile (the program and the content, the requirement of the module, alternative modules, etc.). Then, PASS uses a learning analytics engine to analyze the data by combining both qualitative and quantitative methods. So, it produces reports, feeds, recommendations, suggestions about the modules to be followed for students and tutors.

Another valuable experience is described by Tempelaar, Rienties, Mittelmeier and Nguyen (2018). Their contribution revolves around undergraduate students learning mathematical and statistical methods in a blended environment. During the face-to-face activities, they had problem-based learning (Lotti, 2018) experiences. The online activities, however, consist of Blackboard as the Learning Management System to share the main course information and two external e-tutorials (SOWISO for mathematics and MyStatLab for statistics). These two tutorials provide tests and practical problems for learning. Each phase in the learning path starts with a question; if the student does not
answer the question, he can ask for suggestions to solve the problem or for a fully solved example. In this way, students can choose between two different types of feedback and after receiving them, a new version of the problem is loaded to check the student’s newly acquired skill.

At the New York Institute of Technology, they (Agnihotri & Ott, 2014) developed their own Learning Analytics model to identify at-risk students and to support them during the entire learning process. In their analytics, key risk factors were grades, the student’s certainty in their choice of their major subjects, and financial data. The system was developed to help teachers and tutors determine whether each student was likely to return to their course the following year. In addition, the dashboard gave a confidence measure for the prediction and more precise reasons for the prediction itself. This data provided pointers for discussions with the students about their individual situations and future plans.

In addition, in 2013 some authors (Aljohani & Davis, 2013) presented the possibility of using a device called Quiz My Class Understanding (QMCU) to facilitate the integrated combination of Learning Analytics and formative assessment, giving students constant and immediate feedback. In this scenario, the massive use of Learning Management System (LMS) tools is also included, as they provide constant statistics on the different processes that are involved in learning.

In general, EDM provides the techniques for analyzing data according to the different purposes that the teacher and the educational institution have in mind, providing students (in real or deferred time) the results thanks to which it is possible to rethink learning, to redefine the objectives, to choose a new MOOC rather than a Webinar, an online course rather than a blended one, a formal rather than an informal context, and so on.

5. Discussion and conclusion

The case developed at Purdue University examined in the previous paragraph has been widely discussed and also strongly criticized for various reasons, including the timing and frequency of the signals that the students received and the less that total consensus that these have earned. In particular, on the one hand, the excessive number of instances of positive feedback could make students complacent, and they might end up ignoring them at some point; on the other hand, the negative feedback could reduce the motivation to learn. Numerous studies have been carried out (Arnold, et al, 2010; Caufield, 2013; Straumsheim, 2013; Tanes et al., 2011) aimed at evaluating the efficacy of the project from different perspectives. If, on the one hand, the system presented a series of problems, on the other the signs represented a system which gave students the opportunity to learn how to ask for help at the right time, so avoiding irreparable consequences for their improvement and their academic success.

Despite the numerous criticisms received, the objective is still to design a control system that would allow ongoing assessment of academic achievement. Undeniably, this approach is crucial for the global development of students and digital tools can strongly support teachers and educational institutions in the deployment of this process.

As shown through the additional cases, indeed, the use of LA systems can also be the key to promoting academic integration, inclusion for students at school, and decreasing dropout rates. As well as analyzing the data collected including that obtained through virtual devices, teachers can act in different ways. They can:

• Send personalized emails to students to communicate information about their current performance on a given course;
• Use LA to integrate real-time data about student performance and interaction (with LMS environment) with demographic information and information on their academic background.

The question remains on which data mining techniques to use, on how to navigate the huge amount of data and on what to do with the results obtained. Big Data will be a game-changer in the future of education and the Analytics can be used to design new tools to revolutionize and maybe revitalize teaching.

The now classic concept of Zone of Proximal Development (ZSP) (Vygotskij, 1987) indicates that the development of a student is represented by that area of problem solving skills that can be formed through the help of a more experienced peer or an adult, or that area in which the student does not yet have full autonomy.

As previously argued, Learning Analytics acquires an educational purpose if it used as a tool for personalized learning, and the attention to the formative assessment processes implemented through LA plays a fundamental role in achieving this goal. The systematic observation of those zones of development that represent single individuals and groups (Goos, Galbraith, & Renshaw, 2002) of students can represent an important key to understanding the learning process in order to understand the personalized dimension of the individual and of a group of students. In other words, when choosing LA techniques, reading the results and using them for the realization of a formative assessment, we believe that teachers should ask themselves: what is or what are the students’ ZSPs of which I am analyzing the virtual tracks? What are the ZSPs of the micro or macro-group to be analyzed and supported? What are the ZSPs of the institution in which the didactic intervention takes place?

The analysis of these questions can guarantee the use of an approach that considers learning as a process and formative assessment as a useful tool to support the development of the person as a whole, that makes a connection
between the use of big data and the uniqueness of the student, to propose a learning experience that is truly personalized.

The blending of big data and adaptive technological platforms is represented as a revolution that could modify the education system and process, overcoming the obsolete classroom model, and making real the progressive vision of interest-driven and self-initiated learning.

References


Assessment, King’s College, London.


From Big Data to Learning Analytics for a personalized learning experience
Dipace, Loperfido, Scarinci


Scriven M. (1967), The methodology of evaluation. Lafayette, Ind: Purdue University.


