1. Introduction

The rhetoric about big data in general, and learning analytics in particular, tends to highlight the opportunities and potential benefits that learning analytics might bring to education. Gašević (2018) goes even further suggesting that in the specific case of developing countries learning analytics can: support learning at scale, improve the quality of learning experience, provide personalized feedback to learners, increase numbers of graduates at risk of failure, optimize the use of resources, predict future demands, etc.

But not everybody agrees with that perspective. Boyd and Crawford (2012), among others, argue that this utopian perspective needs to be balanced with dystopian concerns. Boyd and Crawford highlight what they call six provocations: 1) big data changes the definition of knowledge; 2) objectivity and accuracy are misleading; 3) bigger data are not always better data; 4) when the context is taken out, big data loses its meaning; 5) just because data can be accessible, it does not make it ethical; and 6) limited access to big data creates new digital divides.

This report will discuss learning analytics with a special focus on the Latin American reality. This article identifies some of the advances in this field but also highlights what Boyd and Crawford called the emerging digital divide not only between the haves and the have-nots but also between the doers and the do-nots, and between the knowers and the know-nots. The development of learning analytic studies is still considered emergent in Latin America, but there are trends which make us think that this will be a topic of growing relevance in the years to come. In addition to building the necessary technical, financial, academic, and legal infrastructure for learning analytics, it will be relevant to develop and consolidate a dynamic Latin American research network in this field.

This article concludes with some of the challenges that need to be addressed for developing new capacities towards making educational data more actionable in this region.

2. Latin America’s Background and Current Trends

Latin America includes a collection of countries with many similarities. It refers to a vast geographical region that comprises South America, Central America, a part of North America, and the Caribbean. Countries in this region share a common historical and cultural past, but they are highly diverse in many aspects including language, resources, and
educational infrastructure including academic and research centers (Kalergis, Lacerda, Rabinovich, & Rosenstein, 2016).

Latin America is a profoundly socially unequal sub-continent not only in terms of income distribution but also in terms of individual access to public services including education, health, water, and other utilities. The difference in average years of education for adults in the top and bottom income quintiles, for example, ranges from five to nine years in different countries. Available data, which extends back to 1950, suggests that Latin American countries have consistently been among the most unequal throughout the period.

Compared to international standards, much of Latin America can be said to suffer from a massive “secondary school deficit,” with abnormally low proportions of the population achieving secondary education, directly impacting higher education achievement. The most obvious concern perhaps is that as much as three-quarters of the region’s potential labor force possesses at most only a few years of basic primary education. In turn, unequal educational distribution clearly serves as an important channel for perpetuating inequality across generations.

It is fair to mention that there has been some progress at the quantitative level. Over the past two decades, for example, the average years of schooling for Latin America’s adult population (25 and older) increased by 1.7 years (De Ferranti & Ody, 2006). Most Latin American countries are close to achieving universal participation in at least some primary schooling. Earlier gender gaps in school attendance were also narrowed or eliminated over the past decades. However, the substantial improvements in quality indicators have been more difficult to achieve than the quantitative increases in attendance.

Latin American higher education consists of close to 500 million inhabitants in 19 countries. It is important to mention that higher education systems in Latin America need a deep transformation to consistently assure quality in education (higher retention rates, well trained and employable professionals) and science (excellence, international presence, better funding schemes), support smarter diversification, and provide society with the knowledge-based resources needed (Knobel & Bernasconi, 2017).

Internet penetration in the Latin American region is at 59.6%, placing the sub-continent below the worldwide median (“Internet users,” n.d.). This rate is expected to be higher among higher education institutions, where the Internet has played a key role in overcoming the isolation of scientific communities by facilitating exchanges among peers across the world and increasing access to scientific journals.

There is still a long way to go to increase the budget for R&D in order to address and overcome the main challenges that these societies face. Additional efforts are required to build new research centers and train young scientists.

3. Possible Scenarios for an Actionable Learning Analytics

In Latin America, one of the most vulnerable groups is those “out of school and out of work.” Having a growing youth population divorced from activities that allow them to develop new skills and capacities, which affects their employability, not only undermines the future potential of this cohort but could also raise major challenges to society (D’Alessandre, 2013). While enrollment and graduation rates in Latin America increased and dropouts decreased between 1990 and 2010 (Bassi, Busso, & Muñoz, 2015), nearly 10 million Latin Americans between the ages of 15 and 18 are still neither studying nor working (Cárdenas, De Hoyos, & Székely, 2015).
Learning analytics can supply valuable information tools to work on this problem. For instance, it can provide relevant and actionable information by analyzing the impact of learner’s socio-economic context, the school or college’s quality, the learner’s engagement, the effectiveness of the educational systems, among others (see, for example, Park, Denaro, Rodriguez, Smyth, & Warschauer, 2017 or McKay, Miller, & Tritz, 2012). One of the main differences between learning analytics and “traditional” studies of school disengagement is that with the increasing adoption of digital tools (i.e., smartphones, social networks, school management software or online educational resources), which generates an information-rich context, it is possible to have a much more updated (if not real-time) description of the learner’s path. Additionally, proficient deployment of learning analytics can help to identify at a much more granular level when the learners are at risk of leaving the formal education.

As we move into an era of greater usage of online learning, an increasing number of online and blended interactive learning systems are expressing their interest in moving toward higher personalization. Evidence on the effectiveness of personalization is still preliminary (Baker, 2016). Nevertheless, vendors are increasingly offering “personalized” learning systems and analytics. Educational institutions should request evidence on these systems’ effectiveness, as well as transparency on the developed algorithms.

Personalized learning is a popular buzzword symbolizing the potential for data use in education. As Bulger (2016) argues, personalized learning encompasses such a broad range of possibilities – from customized interfaces to adaptive tutors, from student-centered classrooms to learning management systems. Bulger emphasizes that since personalized learning systems are relatively new and largely untested, the impact on students’ regulation of their learning remains unclear, creating tensions between promise and reality.

We argue that moving into the personalization of learning will require additional actions in terms of data privacy. In order to guarantee the quality and integrity of data management as well as user protection, ethical and legal guidelines in accordance with both national legislation and international standards should be followed.

In addition to privacy concerns, it is also necessary to better understand how learners interact with an ecosystem of educational platforms. Considering that more and more learners are learning on several platforms simultaneously (e.g., Moodle, YouTube, WhatsApp, Facebook, Elsevier), it is necessary to conduct analysis across multiple platforms. Several learning analytics studies (e.g., on massive open online courses or MOOCs, Khan Academy, Wikipedia) tend to analyze silos of information (individual online platforms), thereby losing perspective on the multi-platform online user’s behavior.

This more holistic approach, although challenging, can contribute to building a much more comprehensive picture of the learning experience. This is considered a conditio sine qua non before moving towards more ambitious “personalized learning.” As mentioned, adequately addressing ethical, legal, and societal concerns; handling student data responsibly; and adopting policies that protect privacy yet preserve data and ways to link student learning information are essential.

### 4. Effective Models of Learning Analytics for Latin America

Three major adoption models have been identified in learning analytics: predictors and indicators, visualization, and interventions (Brown, 2012; Gašević, Dawson, & Pardo, 2016):

- **Predictors and indicators** include solutions in which data obtained from learning contexts is analyzed, using statistical and data
mining tools, to generate models capable of predicting variables of interest (e.g., performance, student engagement, dropout).

- **Visualization** tools are used to summarize and simplify large amounts of otherwise complex data, thus enabling more effective exploration and interpretation. These are particularly powerful tools for teachers and decision-makers assisting on educational policy formulation.

- **Interventions** concern the derivation of concrete initiatives to shape the learning environment to improve the learning experience.

Effective implementation of the three adoption models is crucial to mobilizing the full potential of learning analytics to tackle endemic problems in the education systems of Latin America such as student dropout, low performance, and disengagement. Predictive models of student dropout are essential to anticipate the problem and create early warnings, giving the education system the opportunity to make timely interventions (Tempelaar, Rienties, & Giesbers, 2015). Addressing different learning needs and interests through personalized learning can help improve the learning experience, thus increasing performance and student retention. Proficient use of learning analytics can support the design of more personalized strategies to detect and address school disengagement (e.g., context-based or personalized recommendations) (Papamitsiou & Economides, 2014).

There are some moderate initiatives towards learning analytics adoption in Latin America. The learning analytics research community in the region reflects what is observed in the international community. On a regular basis, research initiatives are conducted by universities addressing mainly higher education needs (e.g., studies of student behavior in MOOCs). However, the actual learning analytics adoption in the region is still very limited (e.g. limited participation of Latin American proposals during the last conference ‘Learning Analytics & Knowledge 2017’, at the Simon Fraser University, Vancouver, BC, Canada).

Today’s main areas of research in the region are: **performance** (Ferreira, León, Yedra, Gutiérrez, & Ramos, 2015; Manhães, 2015; Costa, dos Santos Silva, de Brito, & do Rêgo, 2015), **engagement** (Santos, Bercht, & Wives, 2015; Santos, Bercht, Wives, & Cazella, 2015) and **dropout** (dos Santos, de Alburqueque Siebra, & Oliveira, 2014; Queiroga, Cechinel, Araújo, & da Costa Bretanha, 2016). Nonetheless, most of the academic production is still at an exploratory stage of “data crunching” and far from real interventions. There is yet a long way to go from academic research to actual learning analytics institutional adoption.

**5. Ethics and Privacy Protection Experiences in Latin America**

Pardo and Siemens (2014) define “personal digital information” as the information about persons captured by any means and then encoded in digital format. In the digital context, Pardo and Siemens (2014) define “ethics” as the systematization of correct and incorrect behavior in virtual spaces according to all stakeholders. They suggest four ethics and privacy principles for learning analytics: “transparency, student control over the data, security, and accountability and assessment” (p. 448).

According to Tobon, (2015) more than half of the countries in the Latin American region have adopted constitutional rights to privacy and/or comprehensive data protection regulation as mechanisms to protect privacy. For illustrative purposes, Table 1 describes the data protection laws and national data protection authorities in the seven most populous countries in Latin America (DLA Piper, 2017).
<table>
<thead>
<tr>
<th>Country</th>
<th>Law/Authority</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>Data protection law</td>
<td>Brazil does not have a single statute establishing a data protection framework. However, the Brazilian Internet Act establishes general principles, rights, and obligations for the use of the Internet. It includes relevant provisions concerning the storage, use, treatment, and disclosure of data collected online.</td>
</tr>
<tr>
<td></td>
<td>National data protection authority</td>
<td>The Brazilian Internet Steering Committee (Comitê Gestor da Internet no Brasil)</td>
</tr>
<tr>
<td></td>
<td>National data protection authority</td>
<td>The National Institute for Access to Information and Personal Data Protection (Instituto Nacional de Acceso a la Información y Protección de Datos Personales) and the Ministry of Economy (Secretaria de Economia)</td>
</tr>
<tr>
<td>Colombia</td>
<td>Data protection law</td>
<td>Law 1581 (2012) contains comprehensive personal data protection regulations. This law is intended to implement the constitutional right to know, update, and rectify personal information gathered in databases or files, as well as other rights, liberties, and constitutional guarantees referred to in the Constitution.</td>
</tr>
<tr>
<td></td>
<td>National data protection authority</td>
<td>Two governmental authorities are designated as data protection authorities: the Superintendency of Industry and Commerce (Superintendencia de Industria y Comercio or SIC) and the Superintendency of Finance (Superintendencia Financiera de Colombia or SFC). The SIC is the data protection authority unless the administrator of the data is a company that performs financial or credit activities under the oversight of the SFC as set forth by applicable law, in which case the SFC will also serve as a data protection authority.</td>
</tr>
<tr>
<td>Argentina</td>
<td>Data protection law</td>
<td>Personal Data Protection Law (25,326) provides much broader protection of personal data closely following Spain’s data protection law.</td>
</tr>
<tr>
<td></td>
<td>National data protection authority</td>
<td>Argentine Personal Data Protection Agency (Dirección Nacional de Protección de Datos Personales)</td>
</tr>
<tr>
<td>Peru</td>
<td>Data protection law</td>
<td>Personal data protection is governed by the Personal Data Protection Law (29733) and the Security Policy on Information Managed by Databanks of Personal Data.</td>
</tr>
<tr>
<td></td>
<td>National data protection authority</td>
<td>The General Agency on Data Protection (Dirección General de Protección de Datos Personales), part of the Ministry of Justice and Human Rights</td>
</tr>
<tr>
<td>Country</td>
<td>Law/Authority</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Venezuela</td>
<td>Data protection law</td>
<td>Venezuela does not have any general legislation regulating data protection. However, there are general principles established in the Constitution.</td>
</tr>
<tr>
<td></td>
<td>National data protection authority</td>
<td>Venezuela does not have a national data protection authority. Various agencies (e.g., the Superintendency of Banks and the National Telecommunications Commission) have data protection authority within their specific jurisdictions.</td>
</tr>
<tr>
<td>Chile</td>
<td>Data protection law</td>
<td>Personal data protection is addressed by several specific laws and other legal authority. There are at least six main laws containing data protection provisions.</td>
</tr>
<tr>
<td></td>
<td>National data protection authority</td>
<td>There is not one regulator who oversees matters relating to data protection. Such matters are resolved by the Chilean courts: The Jueces de Letras (territorial civil jurisdiction), the Appeal Courts (jurisdiction in the first instance in connection with constitutional actions) and the Supreme Court (for cases involving constitutional violations).</td>
</tr>
</tbody>
</table>

Díaz et al. (2015) conclude that in most Latin American countries this kind of personal information is regulated through *personal data protection* laws. Brazil, Colombia, Paraguay, Peru, Argentina, Ecuador, Panama, and Honduras have recognized *habeas data*¹ as a constitutional right. Argentina, Uruguay, Mexico, Peru, Costa Rica, and Colombia have enacted data protection laws based on the European Union Data Protection Directive of 1995. Chile and Paraguay have data protection laws, although they do not have a data protection authority.

6. Potential Barriers to Learning Analytics and Strategies to Overcome Them

The major barriers for learning analytics adoption can be associated with three main components: data, modeling, and transformation (Gašević, 2018). The first one concerns the information on learning activities, which is at the forefront of any learning analytics development. In this regard, data availability and data quality are two fundamental aspects (see, for example, Hazen, Boone, Ezell, & Jones-Farmer, 2014), which oftentimes present huge barriers to learning analytics adoption.

Data availability tends to be less of an issue in higher education since universities often record data on classroom and online courses. In contrast, primary and secondary education institutions frequently lack this kind of data because they do not have the means and resources to access and store it. Uruguay is a rare exception due to Plan Ceibal, a national policy program created to enable technology-enhanced learning in the country (Aguerrebere, Cobo, Gomez, & Mateu, 2017). Plan Ceibal provides a personal device (laptop or tablet) and Internet access to every child and teacher in K-12 education, as well as a comprehensive set of online educational platforms and contents. This governmental agency retrieves a significant volume of data generated from the student’s online activities, creating an invaluable source of information about their learning process. During the last decade, Latin America has turned

---

¹ *Habeas data* is a remedy available to any person whose right to privacy in life, liberty or security is violated or threatened by an unlawful act or omission of a public official or employee, or of a private individual or entity engaged in the gathering, collecting, or storing of data or information regarding the person, family, home, and correspondence of the aggrieved party.
into one of the most proactive regions in the world regarding integration of ICT aimed at social inclusion and the democratization of education systems (Lugo et al., 2016). In most Latin American countries, the telecommunications infrastructure that provides connectivity to educational institutions is decentralized, making it harder to overcome the data availability challenge. That being the case, it is imperative to deal with the legal and technical concerns of the various organizations involved (public and private), and only after these issues have been resolved would it be possible to start the discussion on technical interoperability and multi-platform data collection and integration. Although infrastructure and connectivity in Latin America have improved greatly in the last decade, the Internet penetration rate is still one of the lowest among the regions, making data availability even more challenging.

The second main component of learning analytics adoption concerns models, specifically the importance of developing correct modeling strategies. It has been proven that the “one-size-fits-all” approach does not work for learning analytics, and those models developed for other contexts, while potentially useful, need to be adapted to local realities (Gašević, 2018). It is essential to conduct learning analytics research using “question- and theory-driven approaches” (Gašević, 2018, p. 11) and not just “let data talk.” In this regard, the limited number of experienced learning analytics research groups in the region may constitute an important barrier to field development and adoption. Despite the existence of regional initiatives to develop learning analytics,² with Brazil, Ecuador, Colombia, México, Argentina, and Chile at the forefront, scientific production is still limited (Nunes, 2015), and the connection with practitioners even more so. To mitigate the situation, it is important to, as Gašević (2018) suggests, promote participation in international initiatives (e.g., SoLAR) and cross-institutional collaborations.

Last but not least, an additional challenge of learning analytics is the development of data literacies among different communities, for instance: data generators (those with the skills to collect, select, clean, analyze, produce, visualize, and share quality information) and data consumers (those with the ability to interpret, use, and understand educational data without ignoring its limitations).

We would like to add that in Latin America learning analytics is still a new field for a large sector of the academic community, most likely for policymakers as well. Using Rogers’ taxonomy (2010), we can say that learning analytics is only significant for a limited group of “innovators” and perhaps a few “early adopters.” It is far from being adopted (or even sufficiently acknowledged) by an “early majority,” much less the “late majority” and the “laggards.”

The problem, therefore, is not that the potentials of learning analytics suggested by Gašević (2018) have not been reached yet but rather, that there is a worrying ignorance of the importance of developing a broader awareness and better understanding of learning analytics and related topics across all different groups in our increasingly data-driven society. It is time to discuss at the societal level how to find a balance between learning analytics research, on the one hand, and privacy and data protection issues, on the other, in order to comply with legal regulations as well as with a number of ethical challenges. For this reason, it is of utmost importance to promote the development of new data literacies among decision-makers, academics, and educators as well as key institutions to address the emerging challenges in this field: privacy; informed consent, transparency, location and interpretation of data; data ownership; obligation to act on knowledge (Steiner, Kickmeier-Rust, & Albert, 2015); and algorithmic accountability (Gašević, 2018).

² Examples of Latin American conferences include the Latin-American Conference on Learning Technologies (LACLO); Congresso Brasileiro de Informática da Educação (CBIE); and the Latin American Workshop on Learning Analytics (LALA).
7. Conclusion

Proficient use of available digital information, enhanced by learning analytics techniques, is paramount to support early (re)actions of the different educational stakeholders (policymakers, educators or parents) to the major challenges facing education in Latin America. In the meantime, learning analytics adoption in the region can be considered emergent, where the main focus of studies is yet exploratory with limited intervention experiences. This trend will likely grow in the coming years, leading to broader and more effective adoption. The existing legal framework should be able to guarantee the first level of acknowledgement and regulation on fundamental ethical and privacy matters that emerge with learning analytics adoption (i.e., transparency, student control over the data, and security). Despite progress regarding legal concerns, there are still limitations and principles to be fulfilled when collecting, processing, storing, and analyzing personal data. Additional steps need to be taken, such as having a thorough treatment of ethical matters, pursuing the creation of national ethical committees, and promoting open discussions both regionally and internationally.

In conclusion, it is important to mention that while most of the analysis presented in this paper has been focused on infrastructure (scientific, legal, and technological), we would like to stress the need for a broader-based conversation about learning analytics involving all stakeholders. For this conversation to be fruitful, we need to foster the development of data literacy among different stakeholders while at the same time generating new R&D capabilities and grant programs in Latin America to facilitate the consolidation of a more dynamic academic community in this field.

A key challenge is how educators and other education stakeholders can be involved in the debates around big data to make sure that educational values are also part of how we use data. As we discussed in this paper, there are a number of problems and critical issues about learning analytics that need to be addressed – questions about the quality of data as well as the quantity and the nature of the tools and techniques used. But we also need more transparency to understand how generalizable the results are. Are we being reductive? Are we neglecting aspects of education that are important? This is not a conversation for experts alone. As Selwyn emphasizes (in Centro de Estudios Fundación Ceibal, 2016), an honest, open, and skeptical conversation about data should include everyone involved in education: students, teachers, parents, schools, employers, communities, and private companies. All those directly or indirectly affected should have a say in the way data are used in education.

References


Building capacity for learning analytics in Latin America


ABOUT THE AUTHORS AND EDITORS

Dragan Gašević is a professor and the Sir Tim O’Shea Chair in Learning Analytics and Informatics in the Moray House School of Education and the School of Informatics at the University of Edinburgh, UK. He is the immediate past president and co-founder of the Society for Learning Analytics Research (SoLAR). His research is centered around adoption of learning analytics and learning analytics for self-regulated and collaborative learning. He had the pleasure to serve as founding program chair of the International Conference on Learning Analytics & Knowledge (LAK) in 2011 and 2012, and the Learning Analytics Summer Institute in 2013 and 2014; general chair of LAK’16; and founding editor of the Journal of Learning Analytics. He has co-authored numerous research papers and books, and is a frequent keynote speaker.

Paul Prinsloo is a research professor in Open and Distance Learning in the College of Economic and Management Sciences, University of South Africa. His academic background includes fields as diverse as theology, art history, business management, online learning, and religious studies. He is an established researcher and has published numerous articles in the fields of teaching and learning, student success in distance education contexts, learning analytics, and curriculum development. His current research focuses on the collection, analysis, and use of student data in learning analytics, graduate supervision, and digital identity.

Bodong Chen is an assistant professor in Learning Technologies at the University of Minnesota, USA. His research sits at the intersection of learning sciences, learning analytics, and online learning. He designs learning environments to foster higher-order competencies, examines learning experiences in MOOCs, and develops analytics to facilitate reflective engagement in educational discourse. He has published widely on these topics and is currently serving as co-editor of the Journal of Learning Analytics’ special section on Temporal Analyses. He is co-chair of the membership committee of the International Society of the Learning Sciences.

Yizhou Fan is a PhD candidate at Peking University, China with a major in Educational Technology. His research interests are learning analytics and instructional design in digital learning environments. As a member of the X-Learning Center of Peking University, he participated in the development of several MOOCs and national educational data management systems in China.

Ma. Mercedes T. Rodrigo is a professor at the Department of Information Systems and Computer Science of the Ateneo de Manila University, Philippines. She heads the Ateneo Laboratory for the Learning Sciences, a research group that studies learner behavior and affect when using computer-based educational interventions. Her research interests include technology in education, artificial intelligence in education, and affective computing.
Cristóbal Cobo is the Director of the Center for Research of the Ceibal Foundation, Uruguay and is an associate researcher at the Oxford Internet Institute, University of Oxford, UK. He works at the intersection of the future of education, knowledge, and technologies.

Cecilia Aguerrebere is a researcher at the Ceibal Foundation, Uruguay. Her work is focused on signal processing applied to different domains (image processing, computational imaging, biomedical imaging, and neuroscience) and, more recently, data analysis for learning analytics.

Cher Ping Lim is a chair professor of Learning Technologies and Innovation at the Education University of Hong Kong and is the Digital Learning for Development (DL4D) Network Lead. He is also the Editor-in-Chief of the journal Internet and Higher Education. He was a professor of Education, the Director of International Partnerships, and the Director of the Asia-Pacific Centre of Excellence for Teacher Education and Innovations in Western Australia before moving to Hong Kong in 2010. Over the past two decades, he has engaged with major education stakeholders at the national and international levels – UNESCO, Asian Development Bank, Microsoft, The World Bank, Inter-American Development Bank, Sampoerna Foundation, and various national and local government agencies, among others – as his research and development partners.

Victoria L. Tinio is the Executive Director of the Foundation for Information Technology Education and Development (FIT-ED) and is the DL4D Network Manager. For the past 17 years, she has been involved in the design, development, implementation, and evaluation of a wide range of digital learning programs and projects in schools and alternative learning environments in the Philippines and in the Asian region.