Learning analytics and data-driven education: A growing field

(Analítica del aprendizaje y educación basada en datos: Un campo en expansión)

Daniel Domínguez Figaredo
Universidad Nacional de Educación a Distancia, UNED (Spain)
Justin Reich
Massachusetts Institute of Technology, MIT (USA)
José A. Ruipérez-Valiente
Universidad de Murcia, UMU (Spain)

DOI: http://dx.doi.org/10.5944/ried.23.2.27105

How to reference this article:
Domínguez Figaredo, D., Reich, J., and Ruipérez-Valiente, J. A. (2020). Learning analytics and data-driven education: A growing field. *RIED. Revista Iberoamericana de Educación a Distancia, 23*(2), pp. 33-43. doi: http://dx.doi.org/10.5944/ried.23.2.27105

Abstract

The growing presence of digital mediation systems in most educational spaces —whether face-to-face or not, formalized or open, and at basic or lifelong learning levels— has accelerated the advance of learning analytics and the use of data in education as a common practice. Using digital educational tools facilitates the interaction between students, teachers and learning resources in the digital world, and generates a remarkable volume of data that can be analyzed by applying a variety of methodologies. Thus, research focused on information generated by student activity in digital spaces has risen exponentially. Based on this evidence, this special issue shows a set of studies in the field of data-driven educational research and the field of digital learning, which enriches knowledge about learning processes and management of teaching in digitally mediated spaces.

Keywords: learning analytics; educational technology; data-based education; data science; educational science; educational research.

Resumen

La creciente utilización de sistemas de mediación digital en la mayoría de espacios educativos – ya sean presenciales o no, formales o abiertos, y tanto en el nivel de educación básica como en situaciones de aprendizaje a lo largo de la vida– está acelerando el avance de la analítica del aprendizaje y haciendo que el uso de la información digital sea una práctica común en el campo de la educación. Las herramientas educativas digitales facilitan la interacción entre estudiantes, profesores y recursos de aprendizaje, y generan de manera continua un notable volumen de datos que pueden analizarse aplicando una variedad de metodologías. Esto ha hecho que aumenten exponencialmente las investigaciones que toman como referencia la información que procede de la actividad de los estudiantes en esos espacios digitales. Partiendo de esas evidencias, este número especial muestra un conjunto de estudios en el campo del aprendizaje digital y la investigación educativa basada en datos, que enriquecen el conocimiento sobre los procesos de aprendizaje y la gestión de la enseñanza en espacios mediados digitalmente.

Palabras clave: analítica del aprendizaje; tecnología educativa; educación basada en datos; ciencia de los datos; ciencias de la educación; investigación educativa.

The phenomenon of data-driven education has led to different types of studies. There is a great deal of research using educational data mining that seeks to analyze student behavior patterns and to establish relationships between the variables involved in the learning process and learning outcomes. A second trend refers to studies with a pedagogical approach, which use the aggregated information resulting from the analysis of the data with the aim of improving instructional design, enriching didactic methods and better understanding the role of educational agents. Finally, there is also a significant amount of research that focuses on the institutional derivatives of the use of digital data and seeks to develop frameworks for improving strategic decision-making, organizational design, and curricular policies.

Despite the heterogeneity of approaches and methods in the field of data-driven education, there is a consensus on the need to generate evidence on measuring the skills that students acquire in digital learning situations, how those skills change throughout a course or in lifelong learning situations, and the causal relationships between students' online behaviour and their learning.

A HETEROGENEOUS FIELD

There has been a substantial increase in the amount of research on the impact of digital data on education in recent years. As mentioned before, along with this quantitative growth, the approaches, methodologies and analytical frameworks employed have also diversified. In the first evolutionary stage of the research field, the focus was on the creation and consolidation of a new discipline called learning analytics. But the massive datafication of society has expanded the number of dimensions to be considered around data-based education, opening the field to the social sciences and humanities. Thus, there are now many elements of interest in how data can impact education and, consequently, many possible approaches from an analytical point of view.

Two phenomena are considered representative of this expansion. One is the increasing attention from learning design to educational data sets and data-based algorithms. Historically, educational research has involved several levels, and this is now the case for digital data. Along with the use of available data, there is a growing concern about how to manage this information, questioning who is involved in the process and how the data are aggregated. Data-based educational algorithms are very intrusive elements: they directly influence the practices of educational agents and determine students' learning. Key issues such as cognitive biases, cultural variables, and issues related to previous user skills are embedded in their development (Hartong and Förschler, 2019). Incorporating these elements into the field demands new perspectives that explore not only what can be done with the data, but whether it should be done, how it should be done, and how it fits into existing learning ecosystems (Shibani, Knight, and Buckingham Shum, 2020).

Second, policy and ethical issues are beginning to be a determining factor in data-based educational practices. This moves the focus away from the analytical procedures of a discipline that was initially configured around the aggregation of digital data generated in learning situations. Methodologies from the social sciences, humanities and computing —and their intersections with the arts and natural sciences— introduce new approaches to experimenting with educational practices using digital data (Buckingham Shum, S., Ferguson, R., and Martinez-Maldonado, 2019).

And finally, there are growing concerns about the ethics and potential misuse of data. Questions about ethical risks relate to the ability of data science to influence educational decisions and practices. These fears were also experienced during the evolution of artificial intelligence, and now rise again within the context of education. In order to overcome the criticism of risks, reflections are sliding towards the set of levels where data-based learning systems are used. So, again, comprehensive analytical approaches are proposed, considering the different layers that are part of an educational process (Barocas and Boyd, 2017).

All of this brings attention to diverse areas not initially considered within data-based education, and it also reveals a movement away from the key issues of learning analytics (monitoring assessment, enrolment, dashboards, drop-out, etc.), which are closely related to the aggregation of large volumes of data from learning situations.

CHALLENGES

While the previous decade has bootstrapped research related to the use of data in educational settings, the transference of this research to teaching practices in the classroom or to institutional policies has been rather limited. More than finding solutions, this previous decade has found key challenges (Wilson et al., 2017) and questions that will need to be addressed before we can start reaching the real potential of this multifaceted area that aims to improve education through data. In this section, we frame these challenges in three areas: research, practice and institutional challenges.

The first area that we analyze, research, has been the one that has received the most attention in the field. The number of research papers has risen exponentially thanks to the common use of digitally mediated learning environments that collect large datasets of students' traces. However, most of these studies have been rather inconclusive, focusing on exploratory analyses without clear initial hypotheses or pre-registered experimental design (Wagenmakers et al., 2012). This trend has made it difficult to transform the conclusions of those case studies into a strong foundational base upon which future research should be built. One of the most complex issues has been the systematic mapping of complex learning and cognitive processes into simplistic features that aim to become a representative image of those processes. As these features and models are created based on historical data, one key challenge that emerges is how to deal with the existing intrinsic social biases that are inherently present in those datasets (Holstein and Doroudi, 2019). Most of the reported approaches in the area encompass the use of quantitative research, and a much more limited number of these have actually included qualitative methods or strong connection to cognitive and learning science theory. More efforts will be needed from the community to develop the foundations of the learning analytics field, so that future research can be strongly grounded on top of them. At a more technical level, multiple issues have emerged, such as scaling up algorithmic implementation and interoperability, model transference across contexts or effective feature engineering, among others, that remain to be solved in the coming years (Baker, 2019).

Regarding the challenges in practice, the lack of confirmatory research and grounding in theory, has been perhaps one of the main causes of the low systematic transference from research to practice. It has been very rare to find practitioners applying learning analytics in their classroom that are not also learning analytics researchers. This raises one of the main challenges of transferring results to educators, which is related to data literacy and interpretability (Mandinach and Gummer, 2016). This challenge deals with the assumption that educators already have the necessary literacy to understand the analytics provided by these tools or applications, and thus they should be able to interpret the analytics and act accordingly in a way that can improve learning outcomes. This is partially caused by the fact that many researchers design systems without involving the target user as part of the design process, thus coming up with prototypes that do not fill the necessities or interests of the person that will be using it. Consequently, one of the promising trends in developing learning analytics systems and applications is to apply a strong user-centered design methodology, and implicating the target user in co-design sessions (Dollinger and Lodge, 2018). The K12 and higher education system, which are well known for the high workloads for their educators, will have to think on ways to incentivize and motivate educators to incorporate these practices in their teaching.

Analogous to other educational improvements, the final objective would be that these applications can be seamlessly and systematically embedded within the educational infrastructure, which triggers a number of additional challenges (Tsai and Gasevic, 2017). Institutions will need to integrate these applications within their technical infrastructure instead of developing isolated ad-hoc solutions that will not scale up. These infrastructures will have to introduce privacy options so that learners and educators can opt-in or -out to data sharing in a transparent and comprehensive way (Gursoy et al., 2016). At the same time, administrators will have to implement evaluation programs that can measure the impact of these applications, while heavily taking into account the ethical concerns raised by A/B testing or the potential biases of pseudo-experimental approaches. Perhaps, the biggest challenge will be to align the points of view of the multiple stakeholders involved in learning analytics, which

would include researchers, educators, learners, educational technologists and administrators, among others (Leitner et al., 2019). This necessity might trigger specific learning analytics units or working groups belonging to the institutional units that deal with development of educational innovations.

We see many of these challenges reflected in the articles of the special issue that we present in the next Section. While the previous ten years of learning analytics have been focused on the research side, the next ten years will have to start dealing with the systematic transference to educational settings by focusing on the educators and institutions. Developing learning analytics that are strongly grounded within frameworks and implement user-centered educational applications as well as the orchestration of the different stakeholders involved in the educational process, will hold the key for the transformation of a field of research, to authentic educational impact.

PRESENTATION OF ARTICLES IN THE SPECIAL ISSUE

An interesting aspect reflected in the contributions to this special issue is that the diversity of approaches does not so much affect the purpose of the studies —mostly seeking to improve learning practices—, but rather their formulation. Papers in this volume consist of both theoretical and applied approaches covering a broad spectrum: there are survey studies on the state of the art of data-driven education; there are also studies with quantitative and qualitative approaches, from various fields such as data science, educational data mining and from the field of pedagogy and educational sciences; and finally there are cases focused on knowing more in depth the different learning processes. For a better understanding, the articles have been grouped according to the following topics: contributions focused on the pedagogical component of data analysis, experiences and cases of learning analytics, foresight and trends, and finally papers with reflections on ethics and data privacy and their management in institutional contexts.

In the first contribution, Alyssa Napier, Elizabeth Huttner-Loan, and Justin Reich analyse, in the study entitled "Evaluating transfer of learning from MOOCs to workplaces: A case study from teacher education and launching innovation in schools", the way learning acquired in online experiences can be applied in offline practice. The analysis of the data from a Massive Open Online Course (MOOC) on innovative practices in teaching aimed at school leaders, shows the need to enrich the pedagogical proposals in online professional learning courses to guarantee the involvement of the participants as well as their commitment to transfer the activities beyond the limits of the MOOC and to embed them in their daily work.

The following two papers focus on ways to implement data-based decisions in educational practice contexts. In the article "Data-driven educational algorithms pedagogical framing", Daniel Domínguez proposes a useful framework to guide educators in the introduction of data-based learning systems. Based on the analysis of the design sequence of algorithms and machine learning solutions, he reflects on the need to introduce a pedagogical approach to guide this process in the case of applications aimed at educational contexts. In this way, the essential principles of learning practices are taken into consideration. Continuing in the educational context, José A. Ruipérez-Valiente in "The Implementation process of learning analytics" echoes the scarcity impact of studies on learning analytics in educational institutions and by professionals. He proposes to implement learning analytics through a pragmatic approach that helps to overcome the barriers detected. The process is divided into 5 phases and in its application the close collaboration of all the actors involved is demanded, so as to allow a systematic and productive implementation.

The paper by Gerald Evans and Rafael Hidalgo, "Analytics for Action: An approach to assessing effectiveness and impact of data informed interventions on online modules", also consists of a pedagogical approach to the phenomenon of data. It relates the case of the Open University (UK), where a framework of learning analytics is systematically implemented. The article shows the experience of the teachers responsible for the learning modules in which they have intervened based on the analysis of the digital data recorded. Teachers are considered the key element in data-based interventions to help retention and engagement of students. Findings show that they are satisfied with the training received and the implementation

process, and that there is a variation in the type of measures adopted: from no intervention at all to interventions that extend over multiple iterations.

Below is a set of articles showing experimental analyses from different data mining techniques, predictive models and their transfer to concrete educational situations. The first of the group, by Alexis Gutiérrez, Ángel Manuel Guerrero, Miguel Ángel Conde, and Camino Fernández, entitled "Assessment of students' academic performance based on the analysis of the use of Version Control Systems", focuses on the use of a version control tool in project management, which is also used in educational institutions. The paper analyzes whether monitoring student activity in these tools can predict academic performance. For this purpose, a prediction model was developed and applied in a university degree course. The results conclude that the model predicts the success of students with a high percentage of success.

Also in the field of predictive analysis, in the case of Ignacio Urteaga, Laura Siri, and Guillermo Garófalo, a central topic in the field of learning analytics is analyzed, such as student dropout in online studies. In their article "Early dropout prediction using automatic learning algorithms in online university extension courses", they applied automatic learning techniques to the records of interactions between students and the Moodle learning platform to predict dropout. Specifically, different algorithms were used to generate predictive models and optimize them towards the mitigation of the economic cost caused by dropout. It was found that it is possible to build successful predictive models, and an algorithm based on neural networks was the most suitable.

The paper by Andrea Vázquez-Ingelmo, Francisco José García-Peñalvo, and Roberto Therón, "Benefits of applying the software product line paradigm to generate dashboards in educational contexts", is the last experimental case of this group. It focuses on the design and development of customised dashboards that allow information to be tracked according to pre-established objectives. Dashboards and, in general, the way to visualize and monitor dynamic information coming from digital learning environments is also a classic topic. In this case, the software product line paradigm to generate dashboards is considered suitable for use in educational contexts.

Finally, three papers are included on trends in the field of learning analytics and ethical issues in learning data management. In the field of foresight, the paper by Alejandra Martínez et al. "Achievements and challenges in learning analytics: The view of SNOLA" reviews the state of the art in learning analytics through the work of the Spanish Network for Learning Analysis (SNOLA). The analysis is based on SNOLA archive data and a survey of current network members. Trends and key challenges are shown, including the need for an ethical commitment to data, the need to develop systems that respond to end-user needs and the need to achieve wider institutional impact.

One of the trends identified, ethics, is the focus of the last two articles in the special volume. In the article "Privacy, security and legality in Blockchain-based educational solutions: A systematic literature review", Daniel Amo, Marc Alier, Francisco García-Peñalvo, and David Fonseca conducted a systematic literature review that explored the importance of personal data protection and security in the field of education, through the emerging promises of those interested in using blockchain technology. The results show the importance of understanding the implications of using emerging technologies in education, their relationship with society and current legislation. And directly related, José L. Aznarte shows in "Ethical considerations on the use of data-driven technologies at UNED" how UNED (Spain) is developing an ethical framework for making decisions that involve the use of data from students. It presents a selection of previous references and approaches that have been taken into account, and lists the key questions that should guide educational institutions considering working with student data.

REFERENCES

Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *JEDM, Journal of Educational Data Mining, 11*(1), 1-17. https://doi.org/10.5281/zenodo.3554745

Barocas, S., y Boyd, D. (2017). Engaging the ethics of data science in practice. *Communications of the ACM, 60*(11), 23-25. https://doi.org/10.1145/3144172

Buckingham Shum, S., Ferguson, R., y Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics, 6*(2), 1-9. https://doi.org/10.18608/jla.2019.62.1

Dollinger, M., y Lodge, J. M. (2018). Co-creation strategies for learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 97-101). https://doi.org/10.1145/3170358.3170372

Gursoy, M. E., Inan, A., Nergiz, M. E., y Saygin, Y. (2016). Privacy-preserving learning analytics: challenges and techniques. *IEEE Transactions on Learning technologies, 10*(1), 68-81. https://doi.org/10.1109/TLT.2016.2607747

Holstein, K., y Doroudi, S. (2019). Fairness and equity in learning analytics systems (FairLAK). In *Companion Proceedings of the Ninth International Learning Analytics & Knowledge Conference (LAK 2019)*.

Hartong, S., y Förschler, A. (2019). Opening the black box of data-based school monitoring: Data infrastructures, flows and practices in state education agencies. *Big Data & Society*. https://doi.org/10.1177/2053951719853311

Leitner, P., Ebner, M., y Ebner, M. (2019). Learning Analytics Challenges to Overcome in Higher Education Institutions. In *Utilizing Learning Analytics to Support Study Success* (pp. 91-104). Springer, Cham.

Mandinach, E. B., y Gummer, E. S. (2016). Data literacy for educators: Making it count in teacher preparation and practice. Teachers College Press.

Shibani, A., Knight, S., y Buckingham Shum, S. (2020). Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education, 46*(2020). https://doi.org/10.1016/j.iheduc.2020.100730

Tsai, Y. S., y Gasevic, D. (2017). Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 233-242).

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., y Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*(6), 632-638. https://doi.org/10.1177%2F1745691612463078

Wilson, A., Watson, C., Thompson, T. L., Drew, V., y Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education, 22*(8), 991-1007. https://doi.org/10.1080/13562517.2017.1332026

ACADEMIC AND PROFESSIONAL PROFILE OF THE AUTHORS

Daniel Domínguez Figaredo. Professor and researcher at National Distance Education University (UNED, Spain). His research focuses on digital mediation and theories that support open and connected lifelong learning. In his recent work he has delved into the analysis of data-based open education and knowledge management in digital and mixed environments. He is founding member of the teaching innovation group CO-Lab: Open & Collaborative Laboratory for Teaching Innovation, and Board of Directors of CyberPractices Foundation ID: http://orcid.org/0000-0002-7772-1856 Web: www.daniel-dominguez.com
E-mail: ddominguez@edu.uned.es

Address:
Dpto. Teoría de la Educación y Pedagogía Social
Facultad de Educación–UNED
C/ Juan del Rosal, 14
28040 Madrid (España)

Justin Reich. He is an Assistant Professor of Comparative Media Studies at MIT, and the director of the MIT Teaching Systems Lab, which aspires to design, implement, and research the future of teacher learning.
E-mail: jreich@mit.edu

Address:
Department of Comparative Media Studies
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139, USA

José A. Ruipérez-Valiente. Juan de la Cierva Researcher in the Department of Information and Communications Engineering in the Faculty of Computer Science at the University of Murcia (Spain). He is also an affiliated researcher at the MIT Playful Journey Lab. His research lines are focused on technology-enhanced learning, with a focus on learning analytics and educational data mining.
ID: https://orcid.org/0000-0002-2304-6365 Web: http://joseruiperez.me/
E-mail: jruiperez@um.es

Address:
Departamento de la Ingeniería de la Información y las Comunicaciones
Facultad de Informática – Universidad de Murcia
Calle Campus Universitario, S/N
30100 Murcia, España

Date of receipt: 29/03/2020
Date of acceptance: 29/03/2020
Date of layout: 30/04/2020