Associated conference: “Yes we can!” - Digital Education for Better Futures (EDEN 2023 Annual Conference)

Conference location: Dublin City University (DCU), Dublin, Ireland

Conference date: 18-20 June 2023

How to cite: Prinsloo, P., Slade, S., & Khalil, M. Yes We Can (But Maybe Not Yet): The Impact of Data Poverty on Realising the Potential of Learning Analytics 2023 *Ubiquity Proceedings*, 3(1): 103-109. DOI: https://doi.org/10.5334/uproc.75

Published on: 27 October 2023

YES WE CAN (BUT MAYBE NOT YET): THE IMPACT OF DATA POVERTY ON REALISING THE POTENTIAL OF LEARNING ANALYTICS

Paul Prinsloo, University of South Africa, South Africa

Sharon Slade, Independent Researcher, United Kingdom

Mohammad Khalil, University of Bergen, Norway

***Correspondence**: Sharon Slade: sharon.slade@outlook.com*

Abstract

Learning analytics (LA), as research and practice, has been committed to contributing to understanding students' learning and the contexts in which their learning occurs since its emergence in 2011. Core to learning analytics is the measurement, collection, analysis and use of student data. As higher education becomes increasingly digitised and datafied, institutions have access to not only more data, but also a variety, velocity and granularity of data. However, there is also increasing evidence of data poverty - referring not only to students' unable to afford data, but also institutions without, inter alia, the data infrastructure and analytical skills to realise the potential of LA. In both cases, the fiduciary duty of higher education institutions to provide enabling and supportive learning environments are severely compromised. In this conceptual paper, we provide an introduction to the notion of data poverty as it relates to both students and institutions before mapping key dimensions of student data poverty and their implications for LA. We also discuss institutional data poverty before concluding with a number of pointers to move from 'not yet' to 'yes we can' in providing evidence-led teaching and student support.

Keywords:

Data poverty; Higher education; Learning analytics; student success.

Introduction

Learning analytics (LA) as research focus and practice has promised to impact positively on the quality of teaching and learning. This impact depends on access to students' digital data (amount and quality), by how student learning is understood, and the quality of any subsequent analysis and interpretation (Prinsloo, Slade & Khalil, 2021; Tsai et al. 2020).

What is often less considered are disparities in students' access to sustainable and affordable connectivity, and the impact of this on both the quality and scope of data provided and subsequently on data analyses and institutional responses. Students may become 'invisible' and be disenfranchised, excluded from resource allocation and pedagogical strategies.

In this conceptual paper, we first consider the notion of student 'data poverty' and its impact on effective, appropriate and ethical LA and more generally, on student support and success. In the second half of the paper we consider institutional data poverty and its impact on LA implementation.

Background

To speak of data poverty amid claims of a data revolution (Kitchin, 2014), of how data "is colonising human life" (Couldry & Mejias, 2019), in which "surveillance capitalism" (Zuboff, 2019) celebrates the commercial value of data abundance, and where we all have become our data (Cheney-Lippold, 2017), seems out of place. Despite an apparent abundance of data, some authors point to digital inequalities and a 'big data divide', referring to "those who have access and ownership to large-scale distributed datasets and those that do not" (McCarthy (2016, p. 1132). As such, the data divide raises concerns about students whose right to autonomy and self-realisation is affected due to user profiles created as a result of criteria and data over which they have no control. (Also see Cheney-Lippold, 2017).

Notions of ‘data poor’ and ‘data poverty’ (the sociomaterial effects of being data poor) are more often associated with educational contexts in the Global South (Heeks, 2022; McCarthy, 2016). However, the recent pandemic illustrated that data poverty is a global problem (Budnitz & Tranos, 2022). Many individuals, social groups and communities live off the digital grid, as a result of geographic location, low income or other factors (Graham, 2002). Such individuals may be unable “to get a data contract” (Lucas, Robinson & Treacy, 2020), seriously hampering their access to the affordances of connected services and opportunities in a networked society (Palmer, 2020). In such circumstances, they may also become invisible (Also see Hayes, 2021; Hayes, Connor, Johnson & Jopling, 2022; Palmer, 2016). Such invisibility “affects vulnerable populations within a variety of geopolitical and socio-political contexts, whereby data poverty constitutes a dangerous form of invisibility which perpetuates various forms of inequality” (Milan & Treré, 2020, p. 2).

As well as this, there is evidence that many educational institutions, particularly in (but not limited to) the Global South (Prinsloo & Kaliisa, 2022), do not have the digital infrastructure or capacity to measure, collect and analyse student data. Whilst higher education is increasingly digitised and datafied (Selwyn & Gasevic, 2020), much student and institutional operational data remains in analogue form and is not readily available for analysis. The notion of institutional data poverty can also include institutions who are data rich but are information or analysis poor (Spiker, 2014; Stichter, 2021).

(Student) Data poverty

The importance of access to student digital data for learning analytics is clear. The geopolitical and institutional contexts in which learning analytics have largely been implemented (e.g.,the United States, Europe and Australia) are broadly known for almost universal internet coverage, though coverage does not necessarily equate to access.

For Lucas et al. (2020) data poverty refers to “those individuals, households or communities who cannot afford sufficient, private and secure mobile or broadband data to meet their essential needs” (p. 16). They also state that “Going online is more costly for those who lack digital literacy” (Lucas et al. 2020, p. 9) and “the greater someone’s need for data, the greater the impact reduced access has on their life” (p. 14).

Of particular interest here are Lucas et al’s (2020) seven dimensions of data poverty, namely affordability; choice; infrastructure; privacy and security; quantity; skills; and usability (Table 1 below).

Table 1. Dimensions of data poverty (adopted from Lucas et al., 2020)

Dimension	Attributes
Affordability	Cannot afford to buy data without cutting spending on basic needs such as accommodation, food, etc.
Choice	Lack of access to open market of data services (e.g., barriers to preferential rates offered via contracts)
Infrastructure	Lack of coverage by sufficiently reliable and fast broadband or mobile infrastructure at an affordable price
Privacy and security	Lack of private internet access – it is accessible only via a shared device or public connection.
Quantity	Lack of access to an appropriate amount of data to meet an individual’s essential needs for information (e.g., health, financial, safety, democratic), services (e.g. education), support (e.g., welfare benefits) and social needs (e.g. connection with family, friends and community)
Skills	Lack of digital skills sufficient to access affordable data or understand their data needs.
Usability	Presence of additional usability needs (including language and communication needs, disabilities, and long-term conditions) which demand greater bandwidth and online support.

It is easy to see how these dimensions of data poverty can be applied to students. While affordability of data and its impact on student learning is well-documented (e.g. Mirata, Hirt, Bergamin & van der Westhuizen, 2020), skills

(Ben Youssef, Dhamani & Ragni, 2022); privacy and security concerns (e.g. Prinsloo, Slade & Khalil, 2022); and issues pertaining to choice, infrastructure, quantity and usability are less-researched.

Lucas et al.'s (2020) data poverty dimensions intersect with issues around gender; race; (dis)ability; culture; and socio-economic, political, legal, and technological forces. As a result, students' data poverty which leads to (relative) invisibility may exacerbate inequalities.

Institutional data poverty

Along parallel lines, 'data rich' or 'data poor' institutions are those entities with access to more or less data than others. Companies such as Amazon or Google demonstrate that "Data is more powerful in the presence of other data" (Palmer, 2016, para 3). Palmer (2016) grounded this concept of data richness around time and resources spent "creating centers of excellence around data governance" (para.5) where data governance relates to, e.g., data availability, accessibility, and usability and analytics (Hussain et al., 2018).

Within higher education institutions (HEIs), although student data is plentiful, it is often distributed across departments and in a variety of formats, e.g., data from learning management systems, admission and library systems, creating significant challenges for data collection. Access to student data has also been brought into sharp focus by legislation such as the GDPR, resulting in data access on a need to know basis, as well as many other issues relating to privacy and data security.

To date, few HEIs have the necessary infrastructure to initiate effective data warehouses and to facilitate initiatives in learning analytics (Ferguson et al., 2016; Cechinel et al., 2020), posing barriers to both processing and analysing data at a larger scale. According to Macfadyen et al (2014), HEIs may lack a data-driven mindset, and this lack of available data impacts on LA implementation. As a result, they state that "It may not be surprising, then, that globally, education lags behind all other sectors in harnessing the power of analytics" Macfadyen et al, 2014, p.22).

Student data and institutional capacity can make a difference

Although LA research has been flourishing since 2011, implementation has lagged. In their review of LA in 84 higher education educations in Europe, Tsai et al. (2020), note that implementation is "primarily at small scales and few institutions had a dedicated strategy, policy, or evaluation framework". This echoes the findings of Khalil et al. (2022) who found that implementation remains relatively low and that evidence of impact is also often lacking.

Having said that, there are a number of examples of embedded implementation. The Open University (OU) in the UK has considerable experience in both research and in the application of learning analytics. In 2014, it introduced the world's first policy for the use of student data in learning analytics (Slade and Boroowa, 2014). Since then the OU has moved from small-scale experimentation to large-scale adoption of LA throughout all of its modules and qualifications. In their recent chapter, Boroowa & Herodotou (2022) describe some of the LA work of the OU, including the use of LA to inform the learning design of online courses and predictive learning analytics to identify students at risk. Other institutions, such as Monash University in Australia, have considerable practical expertise built on focused research. For example, its Centre for Learning Analytics has demonstrated capability in LA applications in feedback and assessment; text analytics; multimodal analytics and learning analytics adoption.

In descriptions of what has been needed for successful implementation of learning analytics, a number of recurring themes are identified:

- It is crucial to understand the *purpose* of the use of LA - what specifically is the institution trying to achieve and why? In support of this, having a policy or or adopting a framework such as SHEILA (Tsai et al., 2018) is extremely helpful. Macfadyen et al. (2014) argue that successful institutional adoption of LA requires comprehensive policies that recognise the institution as a complex adaptive system;
- Institutional leadership and management support is vital. Institutions can adopt small scale implementations, but rolling out a wholesale programme of learning analytics needs both support and sufficient human and technological capacity;

- A clear understanding of purpose should lead to clear identification of the data required. Data collection requires planning and integrated systems, as well as robust data protection policies;
- In addition, buy-in from other key stakeholders (students, faculty, support staff, etc) will make a considerable difference to the success of LA implementation. For example Prieto, Rodríguez-Triana, Martínez-Maldonado, Dimitriadis, and Gašević (2019) stress the need to reach a common understanding among stakeholders. Similarly, based on their experiences at the OU, Herodotou, Rienties, Verdin, and Boroowa (2019) recommend providing evidence (of benefits), promoting cross-stakeholder communication, allocating managerial time, and complementing teaching practice.

.In each of these cases, there is a clear message. Successful implementation takes time - LA implementation should be grounded on a clear purpose, have institutional and stakeholder support, and be able to demonstrate its effectiveness in order to become widely accepted. Fundamentally, when institutions have accessible, reliable and complete student data, they are better prepared to realise the benefits of learning analytics.

Conclusion: Moving from 'not yet' to 'yes we can'

Since 2011, LA has promised to deepen our understanding of students' learning and of the contexts in which their learning occurs. Realising this promise depends on a range of variables such as, but not limited to infrastructure, analytic expertise, and integrated and responsive systems, processes and procedures. Data - both access to it, and also its quality and formats - is obviously at the heart of effective learning analytics.

As explored in this paper, student and institutional data poverty is a reality in the context of higher education, and an important factor in realising LA's potential. Student data poverty may result, on the one hand, in invisibility and exclusion from resource allocation and support; and on the other hand, an absence from discussion forums, for example, may result in students being wrongly targeted as at-risk.

Institutions often use data proxies for student learning (e.g., online engagement, logins, and submission of assignments). There is growing awareness that such proxies should be based on research, and be useful and relevant. Whitman (2020) highlights concerns around proxies and 'data doubles', stating that "By making something called behavior and making it legible through data proxies that minimize gaps between data and what they purport to measure, the institution can track students and monitor them" (p.9) but also "The translation of data into predictions is not solely algorithmic; it is also wrapped up in structural inequalities and notions about what society is and ought to be" (p.3). As such, data poverty, in all of its dimensions, flows from, and often deepens existing structural inequalities, or creates new forms of pathogenic vulnerability.

Whilst some HEIs may remain unable to access good quality data internally, there are a number of open source datasets to train models. For example, the OU's learning analytics dataset (OULAD) includes both student demographic data and VLE interaction data (Kuzilek, Hosta & Zdenek, 2017). RapidMiner and Microsoft's Open Education Analytics are other examples of fully open-sourced data integration and analytics frameworks which can be used within the education sector. They include datasets provided by a variety of organisations and educational institutions around the world. Other approaches to obtaining a richer set of data include partnerships with similar educational institutions (see, for example, Krumm, Everson & Neisler (2022)).

As well as data sharing partnerships, educational institutions might consider pairing with other institutions. For example, the not-for-profit Consilience Education Foundation (2023) offers a Learning Analytics Collaborative for schools, forming partnerships between educational researchers, data scientists, and over 100 schools around the world, many in the Global South. In Europe, the Society for Learning Analytics Research (SoLAR) has an established Learning Analytics Community Europe group (LACE) which seeks to "serve as a meeting point for researchers working on LA in Europe, and help identify new opportunities and partnerships for further collaboration and funding" (SoLAR, 2023). It would be a positive move to see active examples of institution-to-institution mentoring to build further capacity and understanding.

Although partnerships can offer a welcome route into access to data, tools and approaches, care must be taken to ensure that shared approaches are contextually appropriate. For example, Krumm et al (2022) cite challenges arising from "misalignments between a partnership's goals and the purposes for which the data were originally collected" (p.26). In the UK, the Centre for Transforming Access and Student Outcomes in Higher Education

(TASO), an educational charity funded by the UK Government, is partnering with Nottingham Trent and Sheffield Hallam universities to evaluate the effectiveness of learning analytics interventions and any dependencies on context and design choices.

In the absence of key resources, educational institutions may opt to outsource their analytics practice. There are clearly advantages in doing this: access to skills, time savings, etc, but these should be balanced with the potential downsides of data security, loss of skills and insight etc.

Learning analytics is a maturing field, and one which thrives best in data rich environments. In realising better futures we should account for how data poverty - whether referring to students or institutions - impacts on the full potential of evidence-led teaching and student support. Evidence suggests that yes, we can - but maybe not yet.

References

Ben Youssef, A., Dahmani, M., & Ragni, L. (2022). ICT use, digital skills and students' academic performance: Exploring the digital divide. *Information*, 13(3), 129. https://doi.org/10.3390/info13030129

Boroowa, A., & Herodotou, C. (2022). Learning Analytics in Open and Distance Higher Education: The Case of the Open University UK. In *Learning Analytics in Open and Distributed Learning: Potential and Challenges*, eds: Prinsloo, P., Slade, S. & Khalil, M.,47-62. Singapore: Springer Nature.

Budnitz, H., & Tranos, E. (2022). Working from home and digital divides: resilience during the pandemic. *Annals of the American Association of Geographers*, 112(4), 893-913.

Cechinel, C., Ochoa, X., Lemos dos Santos, H., Carvalho Nunes, J. B., Rodés, V., & Marques Queiroga, E. (2020). Mapping learning analytics initiatives in Latin America. *British Journal of Educational Technology*, 51(4), 892-914.

Cheney-Lippold, J. (2017). *We are data. Algorithms and the making of our digital selves*. New York, NY: New York University Press.

Consilience Education Foundation (2023). Retrieved from https://www.consiliencelearning.org/

Couldry, N., & Mejias, U.A. (2019). *The cost of connection. How data is colonising human life and appropriating it for capitalism*. Stanford, CA: Stanford University Press.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., & Vuorikari, R. (2016). *Research evidence on the use of learning analytics: Implications for education policy*. Joint Research Centre, Seville, Spain.

Graham, S. (2002). Bridging urban digital divides? Urban polarisation and information and communications technologies (ICTs). *Urban studies*, 39(1), 33-56.

Hayes, S. (2021). *Postdigital positionality. Developing powerful inclusive narratives for learning, teaching, research and policy in higher education*. Brill Publishing.

Hayes, S., Connor, S., Johnson, M., & Jopling, M. (2022). Connecting cross-sector community voices: Data, disadvantage, and postdigital inclusion. *Postdigital Science and Education*, 4(2), 237-246.

Herodotou, C., Rienties, B., Verdin, B., & Boroowa, A. (2019). Predictive learning analytics 'at scale': Towards guidelines to successful implementation in higher education based on the case of the Open University UK. *Journal of Learning Analytics*, 6(1), 85–95.

Heeks, R. (2022). Digital inequality beyond the digital divide: conceptualizing adverse digital incorporation in the global South. *Information Technology for Development*, 28(4), 688-704.

Hussain, S., Akif, C., Tamayo, J. D., & Safdar, A. (2018). Big data and learning analytics model. *International Journal of Computer Sciences and Engineering*, 6(7), 654-663.

Khalil, M., Prinsloo, P., & Slade, S. (2022). *A Comparison of Learning Analytics Frameworks: a Systematic Review*. Paper presented at LAK22: 12th International Learning Analytics and Knowledge Conference (pp. 152-163).

Kitchin, R. (2014). *The data revolution*. London, UK: Sage.

Krumm, A., Everson, H., & Neisler, J. (2022). A Partnership-Based Approach to Operationalizing Learning Behaviours Using Event Data. *Journal of Learning Analytics*, 9(2), 24-37. https://doi.org/10.18608/jla.2022.6751

Kuzilek, J., Hosta, M. & Zdenek, Z. (2017) Open University Learning Analytics dataset. *Scientific Data 4:170171*, Springer Nature. doi: 10.1038/sdata.2017.171.

Lucas, P.J., Robinson, R., & Treacy, L. (2020). What is data poverty? NESTA. Retrieved from https://www.nesta.org.uk/report/what-data-poverty/

Macfadyen, L. P., Dawson, S., Pardo, A., & Gaševic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, 9, 17–28.

McCarthy, M. (2016). The big data divide and its consequences. *Sociology Compass*, 10, 1131-1140. DOI 10.1111/soc4.12436

McCarthy, M. (2022). The big data divide and its consequences. In Carta, S. (ed), *Machine Learning and the City: Applications in Architecture and Urban Design* (pp. 547-559). London, UK: Wiley-Blackwell. DOI:10.1002/9781119815075

Milan, S., & Treré, E. (2020). <? covid19?> The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins. *Social Media+ Society*, 6(3), 2056305120948233.

Mirata, V., Hirt, F., Bergamin, P., & van der Westhuizen, C. (2020). Challenges and contexts in establishing adaptive learning in higher education: findings from a Delphi study. *International Journal of Educational Technology in Higher Education*, 17, 1-25

Palmer, S. (2016, May 1). Rich Data, Poor Data: What the Data Rich Do – That the Data Poor and the Data Middle Class Do Not! Retrieved from https://www.shellypalmer.com/2016/05/rich-data-poor-data-data-rich-data-poor-data-middle-class-not/

Palmer, S. (2020, March 11). Can the data poor survive? Straight Talk. Retrieved from https://straighttalk.hcltech.com/blogs/can-the-data-poor-survive

Prieto, L. P., Rodríguez-Triana, M. J., Martínez-Maldonado, R., Dimitriadis, Y., & Gašević, D. (2019). Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. *Australasian Journal of Educational Technology*, 35(4), 14–33.

Prinsloo, P., & Kaliisa, R. (2022). Learning Analytics on the African Continent: An Emerging Research Focus and Practice. *Journal of Learning Analytics*, 1-18

Prinsloo, P., Slade, S., & Khalil, M. (2021). Learning Analytics: A Primer. Commonwealth of Learning. Retrieved from http://hdl.handle.net/11599/3883

Prinsloo, P., Slade, S., & Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology*, 53(4), 876-893.

Selwyn, N., & Gašević, D. (2020). The datafication of higher education: Discussing the promises and problems. *Teaching in Higher Education*, 25(4), 527-540.

Slade, S., & Boroowa, A. (2014) Policy on Ethical use of Student Data for Learning Analytics. Retrieved from https://help.open.ac.uk/documents/policies/ethical-use-of-student-data/files/22/ethical-use-of-student-data-policy.pdf

SoLAR (2023) Introducing LACE. Retrieved from https://www.solaresearch.org/community/sigs/lace-sig/.

Spiker, S. (2014, November 24). Data rich, analysis poor. Retrieved from https://www.govloop.com/community/blog/data-rich-analysis-poor/

Stichter, A. (2021, June 16). Data rich and information poor. IoT World. Retrieved from https://iiot-world.com/industrial-iot/connected-industry/data-rich-and-information-poor/

Tsai, Y.-S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gaševic, D. (2018). The SHEILA framework: Informing institutional strategies and policy processes of learning analytics. *Journal of Learning Analytics*, 5(3), 5–20.

Tsai, Y., Rates, D., Moreno-Marcos, P.M.,Muñoz-Merino, P.J., Jivet, I., Scheffel, M., Drachsler, H., Delgado Kloos, C., Gašević, D. (2020). Learning analytics in European higher education—Trends and barriers, *Computers & Education*, 155, https://doi.org/10.1016/j.compedu.2020.103933.

Whitman, M. (2020). “We called that a behavior”: The making of institutional data. *Big Data & Society*, 7(1). https://doi.org/10.1177/2053951720932200

Zuboff, S. (2019). *The age of surveillance capitalism. The fight for a human future at the new frontier of power*. New York, NY: Public Affairs.