Tackling the Challenges with Data Access in Learning Analytics Research: A Case Study of Virtual Labs

Kamila Misiejuk [1], Mohammad Khalil [1] and Barbara Wasson[1]

[1] *Centre for the Science of Learning & Technology (SLATE), University of Bergen, Norway*

Abstract

Data plays a crucial role in learning analytics research. The analysis of student digital traces has the potential to not only provide new insights into authentic student learning but also to inform interventions and give students feedback on their performance. However, there are many challenges to implementing learning analytics, such as access to data and difficulties with aligning learning activities with appropriate data collection. This paper describes some of these obstacles encountered during a European project that focused on collaborative learning in virtual labs with help of learning analytics and presents recommendations for future learning analytics initiatives.

Keywords

Learning Analytics, Virtual Labs, Collaborative Learning, Data Accessibility

1. Introduction

Access to data is crucial in the field of learning analytics, which focuses on analysis, sense-making and providing actionable insights from educational digital traces [1, 2]. These digital traces are generated through user interactions while using online learning platforms. Collecting this data is considered to be more cost- and time-efficient than, for example, conducting focus group interviews or surveys [3]. At the same time, digital data on commercial platforms is not collected and stored for research purposes but to support everyday operations, which can result in messy and noisy data [4, 5] or data that is not meaningful or useful for gaining new insights about the learners [6, 7]. Prinsloo et al. [8] raised concerns about the data ecosystem in the field of learning analytics. The authors noted a lack of evidence of a broader understanding of data, insufficient regulations, and a shortage of standardised data generated from various sources.

The accessibility of data for research purposes varies. It can be limited by creating institutional barriers ensuring the authorship of publications for a specific group of researchers [9]. In the case of commercial platforms, data ownership is typically transferred from users to the vendor when registering on a commercial educational platform; however, many commercial vendors share limited data or no data at all with researchers due to a lack of incentives [4, 5]. Furthermore, data is increasingly considered a private asset of companies that can be turned into financial gain [10]. This leads to the monopolization of data ownership [11, 12]. The impact of limited data access on research was reported in other disciplines. For example, Nagaraj et al. [13] showed an increase in higher quality and quantity of scientific output after satellite images of Earth transitioned from a commercial to an open model.

The issues of limited data access for research purposes are common in the learning analytics field but are rarely discussed in the literature. Leitner et al. [14] described the challenges encountered regarding exchanging data with external providers and data ownership in a European project

Proceedings of the Technology-Enhanced Learning in Laboratories workshop (TELL 2023), April 27, 2023, Online.
EMAIL: kamila.misiejuk@uib.no (A. 1); mohammad.khalil@uib.no (A. 2); barbara.wasson@uib.no (A. 3)
ORCID: 0000-0003-0761-8703 (A. 1); 0000-0002-6860-4404 (A. 2); 0000-0003-4897-1394 (A. 3)

involving four universities. Misiejuk [15] reported on the challenges of working with a limited dataset provided by a commercial vendor.

This paper reports on challenges to accessing and collecting educational data for learning analytics in a case study of a European project focused on collaborative learning in virtual labs. In addition, recommendations for future learning analytics initiatives for more collaborative virtual labs are presented.

2. Background

2.1. The importance of data for learning analytics

The underlying assumption in learning analytics is that “generating more information about how learning processes unfold can help us better improve them” [16, p. 120]. To improve and better understand learning processes, educational data is collected, transformed from the raw data into indicators and analysed and, finally, interventions are developed based on the insights [17]. There are many types of educational data that can be captured. The most common is activity data that comes from student interactions in online environments and can include both clickstream data as well as artifacts (e.g., student essays) that students have created. This data can be complemented by demographic, administrative, self-reported or multimodal data (e.g., physiological data or environment data) [18]. Learning indicators are constructed from raw data representing specific behaviors that learners engage in during the learning process [19]. Depending on the types of data captured and the research questions, various methods can be used to analyse the data, such as regression analysis, social network analysis or epistemic network analysis. The value of the learning analytics results lies in their actionability and explainability [16]. Ideally, the interventions informed by the results of the data analysis should happen shortly after the data analysis, improve the learners from which the data was collected, and address a specific issue [20]. Some examples of interventions include dashboards, recommendations, or action suggestions for instructors [21].

The insights from the educational data can be useful for both learners and instructors. Learners can better understand their learning processes and performance, reflect more, and improve their self-regulated learning skills [16]. Furthermore, learning analytics was found helpful in terms of enhancing learner engagement and adaptivity, as well as improving their learning outcomes [22]. Instructors can benefit from learning analytics insights by receiving information about learner behaviour during learning activities, which can help with planning and orchestration [16]. Learning analytics can identify learners that require more support (i.e., struggling learners). Finally, the data can help monitor learner activities in online environments, such as virtual labs, as an instructor naturally would do in a face-to-face class [22].

2.2. Previous research on virtual labs and learning analytics

The pandemic of the Covid-19 has considerably changed teaching and learning in higher education, and laboratories are not an exception. Despite the restrictions of physical contact becoming easier following 2022, post-pandemic practices in education and laboratories have undergone significant changes compared to pre-pandemic times. For virtual laboratories, the exploration has marked a surge in terms of usage, adoption, and research [23]. Although physical interaction with equipment is crucial to get a better sense of experiments, virtual labs may grant students an acceptable level of practical skills they need [24]. For example, students can get a good overall understanding of the practical guidelines, safety procedures, and equipment operation.

From a data ecosystem perspective, virtual labs have generated novel categories of data, creating possibilities to enable learning analytics for further supporting students and optimising the environment of running online labs. Learning analytics is an inter/multi-disciplinary field that has been inspired and guided by diverse disciplines including but not limited to computer science, psychology, and education sciences. The field sustains its definition in 2011 to be the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs [25].

In exploring the potential and challenges of learning analytics in virtual labs, a recent study by Elmoazen et al. [23] provided an overview of this intersection via a systematic literature review. Elmoazen et al. [23] stated that despite the emergence of learning analytics over a decade ago, the number of articles specifically devoted to virtual laboratories remains few. The authors found out that the 21 studies included in their analysis generally remained in a descriptive research stage rather than explanatory and conclusive. For example, learning analytics was utilised to study the perceptions via self-reported feedback [26]. Learning analytics was also used to explain behaviour patterns [27] and assess inquiry-based educational designs by teachers [28]. Elmoazen et al. [23] discussed that while learning analytics should provide actionable insights to relevant stakeholders (i.e., dashboards, collaborative prompts, recommendations etc.), the current state of the field together with virtual labs is still in its early stages. That is, for instance, little is known about dashboards introduced to both students and teachers in virtual labs, collaboration incentives, and insights to lab scientists/facilitators who moderate the virtual experiments.

Despite the fact that collaboration is a key component of virtual laboratories, teamwork incentives are found to be lacking when it comes to learning analytics and virtual labs, hindering further research and development in this area [29, 23].

3. Case study: Learning analytics in the ENVISION_2027 project

The main goal of the European Network for Virtual lab & Interactive SImulated ONline learning 2027 (ENVISION_2027) project is to enhance conventional classroom teaching with virtual laboratories and other digital tools in Bioscience studies. Specifically, the project aims at supporting student engagement and collaboration. In the project, virtual labs were integrated using software from a commercial vendor into multiple biology courses at four partner universities in four countries over three years. The courses were evaluated after each semester and improved for the new iteration. Overall, students reported a positive impact of virtual labs on their learning.

There were some specific goals connected to learning analytics and virtual labs in the project. First, the project aimed to explore ways to implement learning analytics in virtual labs, also real-time learning analytics. Second, learning analytics should have been integrated into the courses to support both student learning and collaborative learning. In addition, learning analytics insights should have informed future course development and teaching.

The commercial virtual lab software, Labster[2], selected for this project features a catalogue of over 300 simulations in multiple disciplines and allows embedding quizzes. The instructors are supported with a teacher dashboard and supplemental resources. Previous studies reported that Labster had a positive effect on student motivation, confidence and learning [30, 31].

It was only possible to obtain data about student attempts at solving a simulation with a timestamp and if they were successful. Student activity data, such as click data, were not shared with the project partners. This made the analysis of data using learning analytics techniques challenging, as it was not fine-grained enough to enhance the virtual lab learning environment. In addition, this data was not detailed enough to inform future course design or teaching. Hence, the evaluation of courses was conducted through student and instructor surveys.

As it was not possible to analyse the data from the virtual lab activities, the focus shifted to the other project goals, i.e. collaborative learning. Typically, virtual lab learning activities in the project were not designed as collaborative tasks. Students were required to solve the simulations as homework individually. In addition, Labster does not support collaboration tasks without the use of third party software [32].

To collect student collaboration data, an open-source instant messaging tool, Discord, was chosen. A recent literature review by Craig and Kay [33] on Discord use in Higher Education reported that it is easy to use and access, as well as increases social presence which can lead to increased student learning outcomes. At the same time, using Discord was found to lead to higher distraction levels. Another study found that students using Discord interacted more, and their discussion was less centralized compared to students using a discussion forum in a learning management system [34].

[2] https://www.labster.com/ (last accessed March 2023)

Integrating Discord with Labster was a challenging task. Communicating while working on a simulation would require students to have both tools open on their devices, providing an inconvenient lab/course structure. Alternatively, they would have to switch between tools leading to distraction. In addition, since the virtual lab simulations were designed to be individual tasks, there was little incentive for students to collaborate while solving them other than help-seeking when they got stuck. As virtual lab activities were assigned as homeworks, students would need to coordinate to meet in Discord at the same time to collaborate. A more natural setting would be for students to collaborate on solving the virtual lab task in a face-to-face setting since most classes take place on campus. However, then there would be no need to use an online messaging tool. This would also require additional data collection through video and audio of student collaboration sessions and would be difficult to scale. Finally, Discord allows for both chat and audio/video conferencing. Text data from a chat is less challenging to analyse, but using video or audio for communication would be more convenient for students while solving a task with another digital tool. Although Discord enables downloading of video and audio using additional software, this would require students to participate in the recording and sharing of their sessions with the researchers. As a result of these constraints, the text data collected from Discord was limited, and student collaboration was assessed through surveys.

4. Discussions and future directions

Digital traces can help discovering more authentic insights into student behaviour in comparison to, for example, surveys that rely on student self-assessment and self-reporting [3]. However, without access to meaningful data, enabling learning analytics is challenging, and thus, new insights cannot be obtained. The issues of data for learning analytics are typically discussed in terms of trust, privacy, and consent [35, 36], equity, diversity, transparency, and inclusion [37, 38, 39], and technical challenges, such as interoperability and integration of multiple data sources [40, 341]. Learning analytics goals cannot be achieved without good quality data of appropriate granularity. One of the main critiques of the field is the focus on "collect[ing] and measur[ing] what is readily available, can be measured and analysed most easily" [42, p. 5]. This case study shows that sometimes this is the only way to conduct learning analytics since researchers have little influence on data quality and granularity from commercial vendors [7] or a long negotiation process with a vendor can lead to a need for solutions to collect the data that may not be grounded in learning theory or be impractical to implement. In this case study, conducting a survey had fewer barriers than conducting learning analytics.

The specific goals for learning analytics implementation in this project faced challenges. The lack of data access limited the analysis options substantially, as student behaviour in the virtual labs could not be analysed. In addition, integration of learning analytics into the courses was not possible since the alternative data collection avenues, such as video and audio, were not scalable. Finally, collaborative learning requires more than "simply placing learners in a group and assigning them a task" [43, p. 1]. The virtual labs in this project were designed to be individual tasks, which resulted in many obstacles in redesigning them to be collaborative learning activities.

This paper describes difficulties and challenges that may be important to consider in other learning analytics initiatives. Although we cannot report on the successful implementation of learning analytics, here are some recommendations based on the experiences from the case study:

1. To map the data available and its granularity with the commercial vendor before deciding on a tool or, if possible, to ask other researchers who worked with a specific vendor about the type and format of data shared.
2. To consider alternative activity data sources if the data from a specific tool is not sufficient, such as multimodal data of students interacting with a tool.
3. To supplement the activity data collection with surveys and other conventional data sources.
4. To align the tools used with the main pedagogical goals, such as supporting collaborative learning, to ensure that alternative data collection scenarios are possible.

5. Acknowledgements

This study is co-funded by the Erasmus+ program of the European Union within the project ENVISION_2027, grant number (2020-1-FI01-KA226-HE- 092653).

6. References

[1] G. Siemens, Learning analytics: The emergence of a discipline, American Behavioral Scientist 57 (2013) 1380–1400. doi:10.1177/0002764213498851.
[2] M. Wilson, K. Scalise, Learning analytics: Negotiating the intersection of measurement technology and information technology, Learning, design, and technology (2016) 1–23. doi:10.1007/978-3-319-17727-4_44-1.
[3] W. Greller, H. Drachsler, Translating learning into numbers: A generic framework for learning analytics, Journal of Educational Technology & Society 15 (2012) 42–57. URL: https://www.jstor.org/stable/jeductechsoci.15.3.42.
[4] C. Fischer, Z. A. Pardos, R. S. Baker, J. J. Williams, P. Smyth, R. Yu, S. Slater, R. Baker, M. Warschauer, Mining big data in education: Affordances and challenges, Review of Research in Education 44 (2020) 130–160. doi:10.3102/0091732x20903304.
[5] R. Kitchin, T. P. Lauriault, Small data in the era of big data, GeoJournal 80 (2014) 463–475. doi:10.1007/s10708-014-9601-7.
[6] J. Roschelle, A. Krumm, Infrastructures for improving learning in information-rich classrooms, in: Measuring and visualizing learning in the information-rich classroom, Routledge, 2015, pp. 3–10. doi:10.4324/9781315777979-7.
[7] A. Krumm, B. Means, M. Bienkowski, Data used in educational data-intensive research, in: Learning Analytics Goes to School, Routledge, 2018, pp. 16–37. doi:10.4324/ 9781315650722-2.
[8] P. Prinsloo, M. Khalil, & S. Slade, S. Learning analytics as data ecology: A tentative proposal. Journal of Computing in Higher Education, 1-29, (2023). doi: 10.1007/s12528-023-09355-4.
[9] C. Pinel, When more data means better results: Abundance and scarcity in research collaborations in epigenetics, Social Science Information 59 (2020) 35–58. doi:10.1177/0539018419895456.
[10] K. Birch, M. Chiappetta, A. Artyushina, The problem of innovation in technoscientific capitalism: data rentiership and the policy implications of turning personal digital data into a private asset, Policy Studies 41 (2020) 468–487. doi: 10.1080/01442872.2020.1748264.
[11] T. Fia, An alternative to data ownership: Managing access to non-personal data through the commons, Global Jurist 21 (2021) 181–210. doi: 10.1515/gj-2020-0034.
[12] G. M. Fulgoni, How limited data access constrains marketing-mix analytical efforts: Why data barriers are preventing marketers from optimizing marketing spend, Journal of Advertising Research 58 (2018) 390–393. doi: 10.2501/JAR-2018-046.
[13] A. Nagaraj, E. Shears, M. de Vaan, Improving data access democratizes and diversifies science, Proceedings of the National Academy of Sciences 117 (2020) 23490–23498. doi:10.1073/pnas.2001682117.
[14] P. Leitner, T. Broos, M. Ebner, Lessons learned when transferring learning analytics interventions across institutions, in: Companion Proceedings of the 8th International Conference on Learning Analytics & Knowledge, 2018, pp. 621–629.
[15] K. Misiejuk, B. Wasson, K. Egelandsdal, Using learning analytics to understand student perceptions of peer feedback, Computers in Human Behavior 117 (2021) 106658. doi:10.1016/j.chb.2020.106658.
[16] A. F. Wise, Learning analytics: Using data-informed decision-making to improve teaching and learning, Contemporary technologies in education: Maximizing student engagement, motivation, and learning (2019) 119–143. doi:10.1007/978-3-319-89680-9_7.
[17] D. Clow, The learning analytics cycle: closing the loop effectively, in: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 2012, pp. 134–138. doi: 10.1145/2330601.2330636.

[18] S. Mu, M. Cui, X. Huang, Multimodal data fusion in learning analytics: A systematic review, Sensors 20 (2020) 6856. doi: 10.3390/s20236856.
[19] A. F. Wise, S. Knight, S. B. Shum, Collaborative learning analytics, International Handbook of Computer-supported Collaborative Learning (2021) 425–443. doi: 10.1007/978-3-030-65291-3_23.
[20] A. F. Wise, S. Knight, X. Ochoa, What makes learning analytics research matter, Journal of Learning Analytics 8 (2021) 1–9. doi:10.18608/jla.2021.7647.
[21] B. Tm. Wong, K. C. Li, A review of learning analytics intervention in higher education (2011–2018), Journal of Computers in Education 7 (2020) 7–28. doi:10.1007/s40692-019-00143-7.
[22] S. K. Banihashem, K. Aliabadi, S. Pourroostaei Ardakani, A. Delaver, M. Nili Ahmadabadi, Learning analytics: A systematic literature review, Interdisciplinary Journal of Virtual Learning in Medical Sciences 9 (2018). doi:10.5812/IJVLMS.63024.
[23] R. Elmoazen, M. Saqr, M. Khalil, & B. Wasson. Learning analytics in virtual laboratories: a systematic literature review of empirical research. Smart Learning Environments, 10(1), 1-20, (2023). doi: 10.1186/s40561-023-00244-y.
[24] J. Glassey, & F.D. Magalhães. Virtual labs–love them or hate them, they are likely to be used more in the future. Education for Chemical Engineers, 33, 76, (2020). doi: 10.1016/j.ece.2020.07.005.
[25] P. Long & G. Siemens. Penetrating the fog: analytics in learning and education. EDUCAUSE Review, 46(5), 31–40, (2011).
[26] H. Considine, A. Nafalski, & M. Milosz. An automated support system in a remote laboratory in the context of online learning. In Proceedings of 23rd International Conference on Interactive Collaborative Learning (ICL2020), Volume 2 (pp. 657–665). Springer International Publishing, (2021). doi:10.1007/978-3-030-68201-9_64.
[27] A. Robles-Gómez, L. Tobarra, R. Pastor, R. Hernández, A. Duque, & J. Cano. Analyzing the Students' Learning within a Container-based Virtual Laboratory for Cybersecurity. In Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality (pp. 275-283), (2019). doi:10.1145/3362789.3362840.
[28] T. de Jong, D. Gillet, M. J. Rodríguez-Triana, t. Hovardas, D. Dikke, R. Doran, ... & Z. C. Zacharia. Understanding teacher design practices for digital inquiry–based science learning: The case of Go-Lab. Educational Technology Research and Development, 69, 417-444, (2021). doi:10.1007/s11423-020-09904-z.
[29] H. Birkeland, M. Khalil, B. Wasson. Understanding Collaboration in Virtual Labs: A Learning Analytics Framework Development. In ICL 2022. Lecture Notes in Networks and Systems, vol 633. Springer, (2023).
[30] S. K. Coleman, C. L. Smith, Evaluating the benefits of virtual training for bioscience students, Higher Education Pedagogies 4 (2019) 287–299. doi:10.1080/23752696.2019.1599689.
[31] M. S. Schäfers, M. Schmiedebach, C. Wegner, Virtuelle Labore im Biologieunterricht, MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung (2020) 140–167. doi:10.21240/mpaed/00/2020.08.31.x.
[32] S. Mathrani, How to use Labster for collaborative learning – 8 tools for your edtech toolkit. 2022, available at: https://www.labster.com/blog/collaborative-learning
[33] C. Craig, R. Kay, Examining the discord application in higher education: A systematic review of the literature, Journal of Digital Life and Learning 2 (2023) 52–66. doi:10.51357/ jdll.v2i2.205.
[34] M. Saqr, S. López-Pernas, Instant or distant: A temporal network tale of two interaction platforms and their influence on collaboration, in: Lecture Notes in Computer Science, Springer International Publishing, 2022, pp. 594–600. doi:10.1007/978-3-031-16290-9_55.
[35] S. Slade, P. Prinsloo, Learning analytics, American Behavioral Scientist 57 (2013) 1510–1529. doi:10.1177/0002764213479366.
[36] O. Adejo, T. Connolly, Learning analytics in a shared-network educational environment: Ethical issues and countermeasures, International Journal of Advanced Computer Science and Applications 8 (2017). doi:10.14569/ijacsa.2017.080404.
[37] M. Khalil, S. Slade, P. Prinsloo, Learning analytics in support of inclusiveness and disabled students: a systematic review, Journal of Computing in Higher Education (2023). doi:10. 1007/s12528-023-09363-4.

[38] T. C. Pargman, C. McGrath, O. Viberg, S. Knight, New vistas on responsible learning analytics, Journal of Learning Analytics 10 (2023) 133–148. doi:10.18608/jla.2023.7781.
[39] A. F. Wise, J. P. Sarmiento, M. B. Jr., Subversive learning analytics, in: LAK21: Proceedings of the 11th International Learning Analytics and Knowledge Conference, ACM, 2021. doi:10.1145/3448139.3448210.
[40] J. Samuelsen, W. Chen, B. Wasson, Integrating multiple data sources for learning analytics—review of literature, Research and Practice in Technology Enhanced Learning 14 (2019). doi:10.1186/s41039-019-0105-4.
[41] M. Y. Amare, S. Šimonová, Learning analytics for higher education: Proposal of big data ingestion architecture, SHS Web of Conferences 92 (2021) 02002. doi:10.1051/shsconf/20219202002.
[42] C. Guzmán-Valenzuela, C. Gómez-González, A. R.-M. Tagle, A. Lorca-Vyhmeister, Learning analytics in higher education: a preponderance of analytics but very little learning?, International Journal of Educational Technology in Higher Education 18 (2021). doi:10. 1186/s41239-021-00258-x.
[43] P. A. Kirschner, G. Erkens, Toward a framework for CSCL research, Educational Psychologist (2013) 1–8. doi:10.1080/00461520.2012.750227.