Bridging Education Learning Analytics and AI: Challenges of the Present and Thoughts for the Future

Laura Hirsto[1,2], Sonsoles López-Pernas[2,1], Mohammed Saqr[2], Teemu Valtonen[1], Erkko Sointu[3] and Sanna Väisänen[1]

[1] *School of Applied Educational Science and Teacher Education, University of Eastern Finland, Joensuu, Finland*
[2] *School of Computing, University of Eastern Finland, Joensuu, Finland*
[3] *School of Educational Sciences and Psychology / Special Education, University of Eastern Finland, Joensuu, Finland*

The Finnish Learning Analytics and Artificial Intelligence in Education Conference (FLAIEC), held in September 2022, was organized by the University of Eastern Finland to provide researchers in learning analytics (LA) and the neighboring fields of education and computer science a platform for sharing their latest research advances, exchanging information and enhancing cross-discipline knowledge, and creating bridges of collaboration. Several important contributions from key figures, researchers and practitioners were presented that have expanded our understanding of current realities, and future challenges.

LA emerged as a field of research in 2011 and has grown exponentially ever since, especially in the context of higher education learning and Massive Open Online Courses (MOOCs) and less on schools or K-12. (cf. [1]). The definition of LA places data and analysis at the center of the mission of the discipline. According to the widely known definition, LA can be defined as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [2]. This implies the need for educational technologies that can generate and collect data from students' participation and interactions during different learning activities. As technology has advanced, so has the integration of technology in learning, which has produced a growing array of data sources. Furthermore, the past few decades have demonstrated an ever-increasing integration of teaching and learning practices in higher education, as well as in primary- and secondary-level classrooms, with a variety of technologies. These technologies—such as personal devices, learning management systems, and cloud services—make classroom learning environments and learning practices interesting venues for LA as well (cf. [3-4]).

The roots of LA research lay in the traditions of educational data mining, where complex statistical methods and algorithms are used to make massive amounts of student data comprehensible and useful for learning support. By virtue of these traditions, LA research can also be seen as having strong roots in the field of computer science, while the role of educational research—the role of actual learning theories—has been less visible. Current investigations have demonstrated the need to further emphasize pedagogy as the key starting point for conducting research and developing learning environments using LA. The purpose of holding a scholarly conference was to highlight the possibilities that educational science and pedagogical perspectives, together with deep and thorough insights from computer science, can bring into learning and pedagogical processes. The themes of the conference ranged from self-regulated learning, learning achievements, and retention in studies, to pedagogical perspectives regarding the use of LA and artificial intelligence (AI) in education. Also, broader themes of learning environments and more thorough investigations of certain learning contexts —such as higher education, vocational education, and elementary schools— were discussed. In addition, more theoretical perspectives as concerns the theory and practice and various uses of LA in educational contexts were included.

Prominent keynote speakers at FLAIEC brought important and topical themes into discussion, providing viewpoints as to the validity of measurements and indicators of LA, and perspectives of LA in collaborative learning processes. Professor **Dragan Gašević** explored the opportunities for, and challenges to, the advancement of validity of measurement in LA, and built on examples from research on self-regulated learning, teamwork, and language learning. Professor **Sanna Järvelä** reflected on the regulation of learning in collaborative contexts and introduced the progress that research on the use of multimodal data can offer. Professor **Dirk Ifenthaler** reviewed the promises and opportunities of LA and tackled the challenges of implementing indicators that generate productive higher education eco-systems. In addition, Professor **Laura Hirsto,** with the OAHOT (Oppimisanalytiikan hyödyntäminen itseohjatun oppimisen tukemisessa koulutuspolun eri vaiheissa) research team, presented in her keynote some key issues of supporting student learning and teachers' teaching with LA, emphasizing the importance of pedagogical perspectives.

Educational science emphasizes the importance of designing learning environments and study modules in pedagogically meaningful ways. This process increasingly requires pedagogical reflection that targets the functions and roles of LA, the kind of data that will be collected, and the way in which the data will be provided for students and teachers [3-5]. Integrating LA, AI, and their potential insights into learning environments and learning management systems (LMSs) can have significant effects on pedagogical processes and students' learning (e.g., [6-7]). However, pedagogical implementation strategies for using LA are found to be challenging, as rich data does not transform easily to support different pedagogical practices [8].

From the viewpoint of computer science, new methodological tools and models using data from various sources are constantly being developed to understand students' learning processes. LA research has evolved from the procedures of educational data mining using the data that different learning platforms have been able to provide. Now, more data and methods that are better pedagogically grounded, based on theories of learning, are needed. Visualizations—provided via dashboards—should be regarded as windows into students' learning processes, serving various pedagogical purposes. Pedagogical actions, tagged to produce data via visualizations, need to be seen as interventions aimed at furthering students' learning and understanding. This calls for a pedagogical design that, along with different technologies, considers the role and place of LA, specifically asking: what are the key processes being pursued that should be made visible via LAs? There is a vast amount of research on the topic of learning, specifically as regards the support of learning online or in the classroom. Still, the visualizations available tend to be based on more traditional models of schooling. Investigating these online and, especially, blended contexts of teaching and learning that target various kinds of student groups will lead to the development of important tools for teachers who are struggling with myriad challenges in dynamic contexts, such that they can better support students' learning in the future.

Earlier research has shown that there are many aspects and indicators that can help us to understand or predict students' achievements or learning processes. Depending on the goals of education or activities of learning, various indicators have been considered important in predicting student achievement. According to Yau and Ifenthaler [9], indicators may predict students' grades, social behavior and engagement, risk of dropping out or of having low performance, or general performance and course completion. In earlier research, indicators such as previous achievement, demographic indicators, and log data from various platforms have been the target of active research; in contrast, curriculum-related indicators such as course characteristics have remained in minor roles [9]. Based on previous studies, the number of indicators affecting students' learning is large and manifold. For example, in the context of personalized and adaptive learning environments, Hemmler and Ifenthaler [10] found over two hundred indicators that were related to learning, either to the learners themselves or to the learning context. In order to take advantage of LA for supporting learning processes, we need to find the most important indicators, i.e., the key pedagogical activities—and find ways to visualize them for students and teachers. It is also clear that different content goals and pedagogical practices benefit from different kinds of indicators; therefore, we need to include teachers as co-designers and engage them more deeply in the inclusion of LA into their pedagogical designs.

Contributions to the FLAIEC Proceedings include ten selected peer-reviewed articles and eight peer-reviewed abstracts. Four of the included articles investigate LA in relation to different pedagogical approaches or models in teaching and learning.

Erkko Sointu, Teemu Valtonen, Sanna Väisänen, and Laura Hirsto, in their article, "Flipped online approach with LA for supporting higher education students' learning: Course feedback results," investigated the use of LA and dispositional LA in the context of a flipped learning approach in online settings in higher education. With dispositional LA, they refer to extending the traditional LA data with data from short questionnaires concerning students' approaches to learning—e.g., emotions related to studying. Based on this data, pedagogical methods and practices were developed during the course to provide adequate support for students, when needed. These students' experiences of the course, which was focused on quantitative research methods for pre-service teachers, were mainly very positive, in terms of interaction between students and teacher, course atmosphere, and instruction in general. However, students assessed their own commitment to studying and their own learning somewhat less positively.

Vesa Paajanen, in his paper “LMS log activity as a predictor of learning success on an undergraduate flipped classroom course of cellular biology,” suggest that monitoring student activity can provide useful information for understanding different student groups' learning preferences and for ways to organize their studying. Paajanen investigated students in a bioscience classroom course conducted using the flipped learning approach. Based on log-data, several challenges were identified in students’ self-regulated learning. Students with high failure risk studied periodically and used significantly less resources and supportive mechanisms that were designed for the course. Paajanen found that high-risk students’ use of, and behavior in, the learning management system already differed in the first week of the course. These results provide teachers with important perspectives on how to identify, at the beginning of a course, the students who will need help in their studies.

Henriikka Vartiainen, Sonsoles López-Pernas, Mohammed Saqr, Juho Kahila, Tuomo Parkki, Matti Tedre, and Teemu Valtonen, in their article, “Mapping students’ temporal pathways in a computational thinking escape room,” explored the applicability of sequence mining and process mining methods in their case study on qualitative video data of a group-based problem-solving situation. Their context was an educational escape room, which combined digital and physical affordances designed to facilitate application of computational thinking. They suggest that sequence mining and process mining methods can be applied to researching collaborative learning through a type of qualitative content analysis.

Finally, Ramy Elmoazen, Mohammed Saqr, Matti Tedre. and Laura Hirsto, in their article “How social interactions kindle productive online problem-based learning: An exploratory study of the temporal dynamics,” explored students’ activity in an online course using problem-based learning. Research was conducted using process and sequence mining approaches in the context of computer-supported collaborative learning. Based on their analyses, students’ most frequent activities were non-argument discussions, followed by sharing knowledge and social interactions. Based on process mining, students’ discussions started with sharing knowledge, followed by evaluating or arguing others’ messages, and ending discussions through social interactions. From the sequence mining perspective, the most common starting activities were social interaction and non-argument interactions. These two approaches seemed to allow researchers to identify different stages of online forum discussions.

Two articles focus on aspects of students’ self-regulated learning. Ji Guo and Guy Trainin, in “Measuring self-regulation: A LA approach,” used confirmatory composite analysis to study the relations between students’ self-regulation, learning behaviors, and academic performance. They suggest that students’ data obtained from a LMS can measure self-regulation and predict performance. According to their analyses, it seems that while many single measures of students’ activities on the LMS were insignificant in predicting students’ performance, self-regulation measures may moderate the effects of driving learning behaviors.

Another self-regulated learning related article, this one by Sami Heikkinen, Sonsoles López-Pernas, Jonna Malmberg, Matti Tedre, and Mohammed Saqr, entitled “How do business students self-regulate their project management learning? A sequence mining study,” investigated the relations between micro-level self-regulated learning processes and academic achievement among business students. The results showed that different kinds of tactics and strategies were used by low- and high-achieving students. They also reflected the importance of understanding the differences among low- and high-achieving students in their micro-level, self-regulated learning processes.

Two papers are related to the use of different kinds of students’ application phase data, and how it could be used for designing, supervising, or counseling interventions in vocational education and training, or to support student selection. Sonsoles López-Pernas, Riina Kleimola, Sanna Väisänen, and Laura Hirsto write about “Early detection of dropout factors in Vocational Education: A large-scale case study from Finland,” and use admission data derived from an institutional admission system in a large vocational school to predict dropout from the initial vocational education and training in a Finnish context. According to their results, students who started a vocational education and training right after basic education seemed to be more likely to drop out compared to older students. In addition, applicants who had a job or were looking for a job seemed to drop out from the studies more easily. The study also showed that in vocational education, large datasets may make it possible for institutions to identify certain risk groups and, perhaps, to tailor proper supervising and counseling practices, where needed.

Mika Nissinen, Elisa Silvennoinen, and Mohammed Saqr, in their paper, “How assessment analytics can help to improve reliability, efficiency, and fairness of entrance examinations,” examined the levels of difficulty, discrimination, and reliability associated with multiple-choice questions, based on a large Finnish law school entrance examination dataset. The result of the analysis by Nissinen et al. suggest that multiple-choice questions may provide a reliable means to identify differences between students and to rank them for acceptance decisions. Success in multiple-choice questions also correlated with achievement in essays and total scores.

One paper focused on the methodological perspectives of text-analysis in order to understand student profiles. In the article, “What are they telling us? Accessible analysis of free text data from a national survey of higher education students,” Sean O’Reilly and Geraldine Gray examine valid and repeatable methods for analyzing survey text responses from large student cohorts. With that, they seek to minimize computational and analyst workload using machine learning. They evaluated clustering and topic modeling to investigate data from a national student survey in Ireland. They suggest that topic modeling provides an effective method to analyze such text data, but it nevertheless requires careful consideration in order to determine the appropriate initial number of topics for configuring the algorithm.

The last reviewed scientific article is a systematic narrative review of LA in the school context. In the article “A systematic narrative review of LA research in K-12 and schools,” Laura Hirsto, Mohammed Saqr, Sonsoles López-Pernas, and Teemu Valtonen suggested that while much of the LA research has focused on higher education, a certain amount of research has also been conducted in elementary-level teaching and learning contexts. Themes that were investigated in K-12 and elementary school contexts included gamification and multimodal methods, and a distinction between research focusing on LA as the target of the study, and LA as the means to study learning, behavior, or, for example, phenomenon such as problem-solving. They also found that most studies lacked a strong theoretical foundation on educational science, which highlights the need to develop more elaborated theoretical bases on school-level LA research.

In addition to the peer-reviewed scientific articles, as this special issue serves as the Proceedings of the First Learning Analytics and Artificial Intelligence in Education Conference (FLAIEC, 2022), eight selected peer-reviewed abstracts are included here. Two of the abstracts refer to the usage of LA in K–12-level schools, and two refer to investigating LA in different pedagogical contexts. Three abstracts are related to the use of LA in the supervising or counseling context. The final abstract brings up an important point regarding a change of perspective—from the aim of understanding students’ learning processes in general, to a more individualized approach toward using LA data in the future.

We, the editors, wish you all to have enjoyable and interesting moments in taking this dive, and as you explore the educational science and pedagogical perspectives of developing and using LA in various educational contexts.

Articles

LA and pedagogical models of teaching

“Flipped online approach with learning analytics for supporting higher education students’ learning: Course feedback results,” by Erkko Sointu, Teemu Valtonen, Sanna Väisänen, and Laura Hirsto.

“LMS log activity as a predictor of learning success on an undergraduate flipped classroom course of cellular biology,” by Vesa Paajanen.

“Mapping students’ temporal pathways in a computational thinking escape room,” by Henriikka Vartiainen, Sonsoles López-Pernas, Mohammed Saqr, Juho Kahila, Tuomo Parkki, Matti Tedre, and Teemu Valtonen.

“How social interactions kindle productive online problem-based learning: An exploratory study of the temporal dynamics,” by Ramy Elmoazen, Mohammed Saqr, Matti Tedre, and Laura Hirsto.

Studying self-regulated learning

“Measuring self-regulation: A learning analytics approach,” by Ji Guo and Guy Trainin.

“How do business students self-regulate their project management learning? A sequence mining study,” by Sami Heikkinen, Sonsoles López-Pernas, Jonna Malmberg, Matti Tedre, and Mohammed Saqr.

Application data in assessing drop-out risk and students’ skills

“Early detection of dropout factors in Vocational Education: A large-scale case study from Finland,” by Sonsoles López-Pernas, Riina Kleimola, Sanna Väisänen, and Laura Hirsto.

“How assessment analytics can help to improve reliability, efficiency, and fairness of entrance examinations,” by Mika Nissinen, Elisa Silvennoinen and Mohammed Saqr.

Methodological approaches to text-analyses

“What are they telling us? Accessible analysis of free text data from a national survey of higher education students,” by Sean O’Reilly and Geraldine Gray.

Review focused on K-12

“A systematic narrative review of learning analytics research in K-12 and schools,” by Laura Hirsto, Mohammed Saqr, Sonsoles López-Pernas, and Teemu Valtonen.

Oral presentations

Pupils and learning analytics in K–12

“Supporting pupils’ reflection with learning analytics during a phenomenon-based study module,” by Teija Paavilainen, Sini Kontkanen, Sanna Väisänen, and Laura Hirsto.

“How do teachers perceive pupils’ use of a learning management system and learning analytics visualizations to support their learning?” by Sanna Väisänen, Laura Hirsto, and Teemu Valtonen.

Learning analytics different pedagogical contexts

“Understanding learners’ needs: Exploratively utilized learning analytics on students' experiences during blended teamwork process,” by Satu Aksovaara and Minna Silvennoinen.

“Game learning analytics: The case of online educational escape rooms,” by Sonsoles López-Pernas, Aldo Gordillo, Enrique Barra, and Mohammed Saqr.

Learning analytics data in advising and supervising

“Implementing a learning analytics dashboard to support academic advising practice: Advisors’ information needs and evaluations,” by Anni Silvola, Jenni Kunnari, Egle Gedrimiene, and Hanni Muukkonen.

“A Chatbot-Guided Learning Experience in The Inquiry Science Classroom,” by Jennifer Davis.

"Learning analytics in Moroccan higher education: Justifications for use and challenges for successful implementation," by Abdelkhalek Zine and Abdelali Kaaouachi.

Methodological approaches

"The idiographic paradigm shift needed: Bringing the person back into research and practice," by Mohammed Saqr and Sonsoles López-Pernas.

Acknowledgements

This research was funded by the Business Finland through the European Regional Development Fund (ERDF) project "Utilization of learning analytics in the various educational levels for supporting self-regulated learning (OAHOT)" (Grant no 5145/31/2019). We would like to thank our funder.

References

[1] Lee, L. K., & Cheung, S. K. (2020). Learning analytics: Current trends and innovative practices. Journal of Computers in Education, 7(1), 1–6.

[2] Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review,46(5), 31–40.

[3] Valtonen, T., Hirsto, L., Sointu, E. & Väisänen, S. (2022). Learning analytics pedagogy—possibilities and challenges. In T. Bastiaens (Ed.), Proceedings of EdMedia + Innovate Learning (pp. 362–366). Online: Association for the Advancement of Computing in Education (AACE). Retrieved January 27, 2023 from https://www.learntechlib.org/p/221313

[4] Väisänen, S., Hallberg, S., Valtonen, T., Tervo, I.-A., Kankaanpää, J., Sointu, E., & Hirsto, L. (2022). Pupils' experiences of learning analytics visualizations in supporting self-regulated learning in an elementary school classroom. Seminar.Net, 18(1). https://doi.org/10.7577/seminar.4690

[5] Sointu, E., Valtonen, T., Hallberg , S., Kankaanpää , J., Väisänen , S., Heikkinen , L., Saqr , M., Tuominen , V., & Hirsto, L. (2022). Learning analytics and flipped learning in online teaching for supporting preservice teachers' learning of quantitative research methods. Seminar.Net, 18(1). https://doi.org/10.7577/seminar.4686

[6] Hirsto, L., Sointu, E., Valtonen, T., Turtiainen, M. & Väisänen, S. (2022). Learning analytics in teaching and learning processes in multiple contexts. In T. Bastiaens (Ed.), Proceedings of EdMedia + Innovate Learning (pp. 359–361). Online: Association for the Advancement of Computing in Education (AACE). Retrieved January 27, 2023 from https://www.learntechlib.org/primary/p/221312/

[7] Hirsto, L., Valtonen, T., Saqr, M., Hallberg, S., Sointu, E., Kankaanpää, J. & Väisänen, S. (2022). Pupils' experiences of utilizing learning analytics to support self-regulated learning in two phenomenon-based study modules. In E. Langran (Ed.), Proceedings of Society for Information Technology & Teacher Education International Conference (pp. 1682–1688). San Diego, CA: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/primary/p/220967/

[8] Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. Educational Technology & Society, 15(3), 42–57.

[9] Yau, J., & Ifenthaler, D. (2020). Reflections on different learning analytics indicators for supporting study success. International Journal of Learning Analytics and Artificial Intelligence for Education, 2(2), 4–23. https://doi.org/10.3991/ijai.v2i2.15639

[10] Hemmler, Y., & Ifenthaler, D. (2022). Indicators of the learning context for supporting personalized and adaptive learning environments [Paper presentation]. 22th IEEE International Conference on Advanced Learning Technologies (ICALT 2022), Bucharest, Romania, January 7, 2022.