Learning analytics: prospects and challenges

Laslo Šereš
University of Novi Sad, Faculty of Economics in Subotica, Subotica, Serbia
https://orcid.org/0000-0002-3435-2579

Veselin Pavlićević
University of Novi Sad, Faculty of Economics in Subotica, Subotica, Serbia
https://orcid.org/0000-0002-4690-3040

Goran Petrović
Institute of Field and Vegetable Crops, Novi Sad, Serbia
https://orcid.org/0000-0002-0287-0881

Danijel Horvat
University of Novi Sad, Faculty of Economics in Subotica, Subotica, Serbia
https://orcid.org/0000-0003-1166-2125

Rajko Ivanišević
University of Novi Sad, Faculty of Economics in Subotica, Subotica, Serbia
https://orcid.org/0000-0002-0601-2884

Abstract
Owing to its high promises for improving learning support, teaching, and learning outcomes in higher education, learning analytics has captured much interest from both academics and practitioners over the last several years. Considering that it is rooted in several disciplines, researchers and practitioners have approached learning analytics from a range of perspectives. Although many studies concerning learning analytics have highlighted its great potential for improving learning practice, there is little evidence of successful transfer of the suggested potential into the practice of higher education happening. This clearly indicates a need for rethinking many aspects of learning analytics usage: first, the goals that can be achieved, but also the actions necessary to attain these goals. The aim of the descriptive research presented in this paper is to provide an updated and realistic view of the state of the art in learning analytics, its potential benefits, and tangible challenges that need to be overcome for a successful application of learning analytics as educational technology.

Keywords
Learning analytics, technology-enhanced learning, action analytics, learning

Introduction

Contemporary teaching and learning practices are considerably influenced by the integration of digital technology into higher education. Data, mainly available from online learning environments, can be very useful in improvement of students' learning. Empowering interactions and communications within a virtual environment (Broadbent & Poon, 2015), online learning became an integral part of higher education, where there is an obvious need to shift its focus "from providing access to university education to increasing its quality" (Lee, 2017). Learning Analytics (LA) systems are implemented by the higher education (HE) institutions in order to improve their understanding and support of student learning (Schumacher & Ifenthaler, 2018). LA emerges as a fast-growing and multi-disciplinary area of Technology-Enhanced Learning (Ferguson, 2012),

which forms its own domain (Strang, 2016) with an evolving interest among practitioners and researchers.

The primary purpose of learning analytics is to improve learning, which is achieved through analysis and representation of data concerning learners and learning environments. It provides teachers with a new lens through which they can better understand and advance the learning process. In LA, information about learners and learning environments is used to "access, elicit, and analyse them for modelling, prediction, and optimization of learning processes" (Mah, 2016). The emergence of LA is closely linked to the impressive increase in the amount of data available on learners, as well as with management approaches focused on quantitative metrics – often inconsistent with the educational perspectives of teaching. However, there is a belief that LA can contribute to a better understanding of students and more efficient use of limited resources.

Among many definitions of learning analytics, the most popular is certainly the one adopted at the first International Conference on Learning Analytics held in 2011: "Learning analytics is 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."

Analyzing the previous definition it is easy to notice that it includes three basic elements.

Data used in LA is usually gathered while the learning process is in progress and concern information about learning environment, learning interactions, learning outcomes and, quite naturally, the students. The typical data sources for learning analytics are:

- Student Information Systems (SIS), as a source of academic and demographic data
- Learning Management Systems (LMS), as a source of data about students' activities and performance
- Other systems as a source of various types of information, which could incorporate library records, consumption patterns of electronic learning material, data on social network interactions, etc.

Data analysis refers to the process of gaining actionable insights from the collected data (Fig. 1). The analysis is based on machine learning techniques utilizing various mathematical and statistical algorithms. It can be said that more sophisticated algorithms usually lead to more

valuable insights, but at the same time these algorithms also set significantly higher requirements in terms of volume, type, time frame and other characteristics of data. Therefore, in the implementation of the learning analytics process, the greatest skill is to choose appropriate data and algorithms.
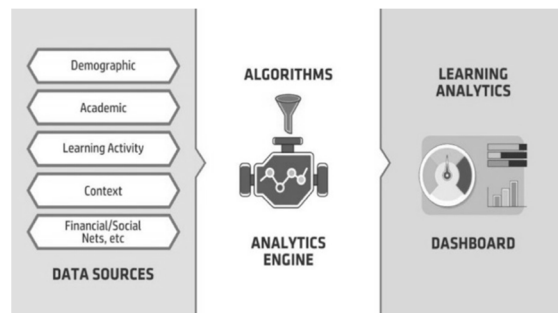

Figure 1 Learning analytics as a process of obtaining insights from the collected data
Source: Omedes, 2018

For the results of the analysis to be in the instrumental to improvement, they must be followed by appropriate actions. Absence of appropriate action can be considered a complete failure and render the entire LA process meaningless. The quality of predictions can be considered irrelevant if one is unwilling or unable to turn them into appropriate action. Although this is usually understood, it should be emphasised that the appropriate internal processes need to be arranged for interventions to happen.

In many fields of activity, especially new ones, there are difficulties in making clear distinctions between related efforts, as well as different interpretations. When it comes to learning analytics, it is not difficult to notice the existence of certain overlaps with two other emerging areas: academic analytics and educational data mining (EDM). The focus of academic analytics is not on individual students and courses, but rather on the institutional and national level (Long & Siemens, 2011). The main subject of interest of the educational data mining (EDM) is the development of methods for the analysis of educational data, where it is much more focused on technical challenges than on pedagogical issues (Ferguson, 2012). Unlike EDM, learning analytics is primarily about learning, more specifically, about the generation of actionable (learning) intelligence related to the possibility of using insights gathered from data in order to improve learning (Campbell, De Blois & Oblinger, 2007).

Campbell, De Blois and Oblinger (2007) set out five steps of the LA process: Capturing, Reporting, Predicting, Acting, Refining. These five steps occupy a central place in the learning analytics cycle presented by Clow (2012) (Fig. 2). The cycle starts with students, whose main characteristic is the generation of data. Students can be either students in a traditional higher education environment or attendees in a less formal context. Data can refer to demographic information, network activities, estimates, and other types of data. These data are processed and converted into metrics whose values serve as the basis for taking actions that affect the participants. The content and way of visualizing metrics is very diverse: from a simple monitoring of learning progress, to a more complex comparison of achieved results with the desired (reference) values or a graphical presentation (visualization) of activity in an online forum. Actions are also very diverse and can be initiated by different subjects: actions can be taken by students in response to metrics that compare their activity to that of their colleagues, but also by teachers who contact those students for whom the need for additional support has been identified.
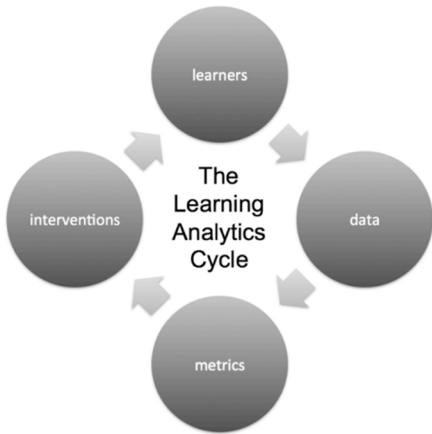

Figure 2 The Learning Analytics Cycle
Source: Clow, 2012

There is no doubt that learning analytics has a number of drivers and facilitators. Like many other areas, there is an obvious pressure towards the implementation of performance management, metrics and quantification in the field of learning. On the other hand, an increasing amount of data on students and learning is available, as more and more learning takes place online. In theory, every page visited, every click made can be easily memorized. Third, advances in big data have led to the wide availability of statistical and computational tools needed for managing large data sets (Sønderlund, Hughes, & Smith, 2019).

As an emerging and promising field, LA is attracting the interest of a growing research community. There are now conferences and special issues dedicated to this topic, as well as established dedicated international research network (Society for Learning Analytics Research – SoLAR). Along with growing interest, vendors of learning technology have been offering an increasing number of learning analytics packages, which is evidenced by the fact that learning analytics and its related technologies have been a part of the "Horizon Report" for a while.

2. Learning analytics as it is

Learning analytics cannot be considered as an established academic discipline with well-defined methodological approaches, but rather a field of enquiry and a random selection of promising techniques, tools, and methodologies. Although this may seem like a strength that allows rapid development, this lack of coherent epistemology essentially appears much more as a weakness and an obstacle to further development (Ferguson, Clow, Griffiths & Brasher, 2019).

Indeed, vendors of learning technology are providing an increasing number of analytics packages. Different packages provide different levels of sophistication in terms of data analysis. Most of the existing solutions are providing only descriptive learning analytics. These types of solutions provide an understanding of the past but do not influence the present or provide any additional help in terms of getting insights into the future events. Unlike these solutions that tend to be reactive, lately we are witnessing a noticeable shift towards predictive learning analytics which should be proactive by influencing the present and thus improving ongoing learning processes. The transition from descriptive to predictive analytics is not easy at all: first of all, it requires accurate and readily available data (which many organizations really do not have) and appropriate algorithms. Below, we describe some of the methods and techniques that currently attract the most attention when it comes to applying LA in practice.

As a result of the aspiration to move from descriptive to predictive analytics, predictive modelling has become one of the most current topics in LA. Application of predictive modelling in education offers many possibilities. Predictive modelling is often used to estimate the probability with which an individual student will complete a course, with the purpose of providing focused support to certain students in order to improve the

overall completion rate. Based on large dataset containing information about previous students who took the course and sophisticated mathematical techniques a model is developed to be applied it in the future to the information available for current students. The purpose of applying the model to the data on current students is to perform a quantified assessment of the course completion probability for each student. The results of this prediction can be presented to teachers, department heads, administrators, and others in a certain form (usually within a dashboard-type control panel). In principle, there is a certain similarity between predictive modelling and a teacher noticing which students have difficulties in learning, and providing them with additional help based on their observations; predictive modelling could be seen as a kind of extension of this ability to the world of online learning. However, it is necessary to note that there are some important practical differences: in learning analytics, insight is not restricted to student's teachers and can be used directly, without involving teachers at all, to initiate actions and interventions. If we want to be realistic, it should be noted that these models are not (always) perfect. If the probabilities created by these models are not completely accurate, then estimates of the student's chances of completing a course based on available data cannot be reliable. However, practice shows that these models are still much more often right than wrong, and, with a certain amount of caution, they can be used to improve student completion rates.

When speaking about Social Network Analysis (SNA), it is necessary to point out that there are specific tools specially developed for the context of online learning, such as Social Networks that adapt pedagogical practice - SNAPP (Bakharia & Dawson, 2011). SNAPP enables monitoring of students' activities on LMS/VLE forums, showing a diagram of the social network which graphically depicts the number and strength of connections between students. SNAPP significantly facilitates the identification of students who are completely excluded from the network or students who are central to the network (and who can be treated as key drivers of communication). It can also be used to determine the pattern of interaction on the forum as well as for identification of stand-alone (isolated) groups that interact with each other, but not with those outside the group.

There are examples that show that SNA can find its successful application in more complex educational contexts. We could mention here the research conducted by Suthers and Chu (2012) in which they, based on SNA, researched the community for professionals in education "Tapped In" (http://tappedin.org). Instead of simple diagrams of social networks, they opted for a more detailed and richer approach, based on an "associogram". They were able to identify real communities by relying exclusively on data describing online activity on the site, without taking into account any other data (considering affiliation, geographical location, etc.). The identification of communities among the student population, provided by this approach, can be of great importance when it comes to making decisions regarding placements, group work, projects and other things.

A common feature of the examples given so far is their predominant reliance on quantitative data created by learners. However, today it is equally possible to analyse qualitative data, primarily owing to advances in computing that are clearly reflected in areas of natural language processing and latent semantic analysis. The analysis of textual data has long transcended simple frequency counts and is nowadays done in a richer, more meaningful way. Content analysis and semantic analysis can be very useful in determining students' contributions to an online forum and the extent to which their online talk is exploratory, as well as in offering suggestions how they might contribute more effectively.

A recommender (or recommendation) system (platform or engine), is an information filtering system which is used to predict items (or ratings for items) that the user may have an interest in. These systems are used in a variety of areas, with commonly recognised examples taking the form of playlist generators for video and music services (Netflix, YouTube), product recommenders for online stores (Amazon), or content recommenders for social media platforms (Facebook, Twitter) and open web content recommenders. The techniques on which these systems are based can also be applied in the educational environment. The recommendation system could make suggestions to the student regarding learning resources, primarily based on what resources where previously used and the rating of their usefulness, but also the experiences of other students. In the case of conventional universities with inflexible curriculum, this approach would have limited effects: in such settings, students usually have limited options for choosing the track of their

study. Along with the increase in the level of studies and the change in the character of studies, which at higher levels is increasingly taking the form of research, the application of this approach could have much greater potential. However, this approach could, as expected, be most widely used in various forms of open and less formal learning.

The examples given above illustrate the potential benefits of applying learning analytics. However, the application of learning analytics inevitably leads to a number of issues. The first and probably the most difficult issue concerns the ethics of personal data. In most countries there is comprehensive data protection legislation, while universities usually have their own data governance policies. In cases where learning analytics is applied within an explicit research context, all activities are carried out under the supervision of an appropriate body (ethics committee or audit committee). However, in cases where learning analytics is applied outside this context, all responsibility regarding ethical standards rests with practitioners.

One of the main drivers of learning analytics' research and application is the expectation that it will improve students' learning outcomes. However, according to comprehensive research conducted by (Viberg, Hatakka, Bälter, & Mavroudi, 2018) this expectation has been confirmed by only few studies (9%) (Fig. 3).
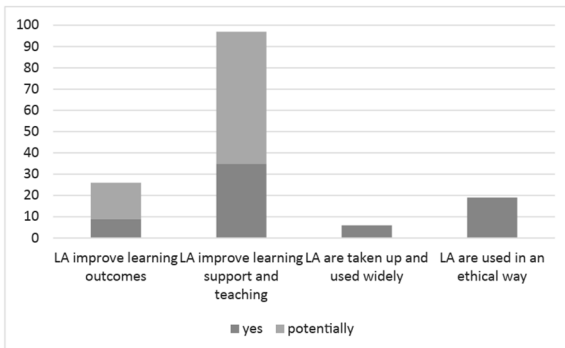

Figure 3 Evidence for learning analytics in higher education (%)
Source: Viberg et al., 2018

As shown in Fig. 3, research indicates that there is a belief in the high potential of learning analytics in terms of improving learning support and teaching, as well as improving learning outcomes, but over the years we have not witnessed many examples of successful transfer of this potential to higher education. This raises many questions and the need to rethink how the potential of LA could be better exploited in higher education practice.

Below we will try to highlight some of the barriers and possible causes of this situation:

- There is no doubt that learning analytics is based on data, but at the same time it is often criticized for being rather theoretical, more precisely, for not being explicit enough in terms of its theoretical basis. There have been several attempts to establish these theoretical foundations by various authors (Clow, 2012; Dawson 2008; Atkisson & Wiley 2011), but none of the offered solutions are universal, and there is an ongoing risk of treating the gathered data as the data that matters. One must not overlook that the choice of what is measured, the choice of metrics, is equally critical when it comes to learning analytics. It is very important for education system to be set up to optimize metrics that really concern learning, otherwise it is not realistic to expect a progress in learning optimization.
- The application of LA in practice is associated with the use of purpose-developed tools and solutions, but being new and popular, also means that the term "learning analytics" is used by the industry in a variety of ways. Unfortunately, some of the offered solutions do not reflect the spirit and/or are not in line with the technological progress in the field of LA.
- There is a tension between shaping education as an economic activity and as an activity that primarily concerns the effective acquisition of knowledge and competencies. This tension has very practical consequences for teachers and students that are manifested in limited resources, class sizes, and general time pressures. The application of quantitative metrics for measuring teachers' practices in such a resource-constrained environment makes teachers very vulnerable in terms of accountability processes.

3. How to increase the odds of success in LA endeavours

In analytics, and thus in learning analytics, data is treated as a primary raw material and has the status of a very sensitive asset. The correct approach in collecting LA data can be described as an effort to gather as much useful data as possible and as little sensitive as required.

Although some tend to characterize learning analytics as "big data in education," a typical learning analytics project does not involve big data as such. The term "big data" refers to data that is so large, rapidly accumulating, or complex that it is difficult or impossible to process using traditional methods. Some of the characteristics of these data (known as the Vs of big data) may not always apply to the data collected in a learning environment. Therefore, there is often no need for the real Big Data processing associated with specific computing tools that can handle the complexity of the 4 Vs. For the majority of LA projects data analysis techniques that are commonly used for mid/small data and concern standard machine learning and analytical algorithms are quite enough.

Despite many promising descriptions of LA, its consideration should always start from the following facts:

- The world around us is increasingly data-driven – it is not realistic to expect education to be an exception
- Learning analytics by improving education make educational institutions more competitive – don't forget that LA is useful only if its implementation results in coherent action.
- To really benefit from LA as an important tool of increasing student success the institutions must implement predictive analytics – descriptive analytics is useful but not enough.

Implementing learning analytics projects in HE may be challenging (Francis, Broughan, Foster & Wilson, 2020). This is especially true for an institution's first LA project, the success of which may be crucial to the attitude towards further LA efforts. In order for an institution to be successful in implementing LA projects (especially the first one), there are certain recommendations that should not be ignored:

- Start modestly: The initial LA project should have a limited scope and duration. It is not a good practice to start with a project that covers the entire company.
- Focus on the problem, not the technology: Technology is useful to the extent that it helps to solve a problem, consequently higher education institutions should focus primarily on specific problems they want to solve when launching LA projects/solutions.

- Successes on smaller projects open the door to larger ones: Proving on (simpler) examples that LA can improve current learning processes is an effective approach to gaining trust and valuable experience.
- Involve all the critical stakeholders from the very beginning: This will help to ensure their long-term support.

Following a more formal approach to learning analytics implementation can also contribute to the greater success of this process. There are several learning analytics readiness assessment tools available to HEIs to conduct self-assessment on how they score in relation to each of the elements critical to an endeavour to establish an analytics-driven culture. There are several elements that can be highlighted as the most significant:

- Data capture: Learning analytics only considers events that leave a digital trace, but it is evident that the digital environment is not the only place where students interact. It means that learning analytics is not capable of providing a holistic view of the entire educational environment.
- Data variety: Learning analytics is much more than a simple extension of LMS. Learning analytics should combine data from different sources, which sometimes appears to be a great challenge.
- Comparable analytics: Currently, no open standard defines relevant metrics and their significance. This represents a significant obstacle when it comes to comparative analyses in which standardized learning analytics metrics should be compared.
- Prediction accuracy: All statistical conclusions are subject to some sort of error. Depending on its expected size and practical importance, we assess whether this error is acceptable or not. And in the case of learning analytics, as for any other analytical process, it can hardly be expected to be (ever) perfect.
- Partial view: Learning processes are much more complex than they seem at first glance and how they are usually treated. We must not forget that we are dealing with people who cannot be fully defined by equations, at least for now.
- Data literacy: Consumers of analytics need special skills to properly interpret the results and apply adequate actions.

Although all the above present reasonable challenges, ethics should probably be highlighted as the most relevant of all. Respect for ethical principles must be treated as a fundamental requirement in all phases of the learning analysis process. We should never forget that learning analytics essentially predicts human success or failure. Therefore, it is very important, or even critical, how we obtain the data and predictions, and even more how we act in relation to these predictions. It is definitely necessary to be highly sensitive to ethical issues.

Conclusion

Learning Management Systems are able to provide information (statistics) on average grades, student progress, time spent on online learning platforms, etc. However, learning analytics is not restricted to data provided by the LMS, it goes far beyond the LMS: in LA we are trying to combine data extracted from the LMS with data available in other systems to obtain more relevant insights. We are making a shift from computing simple statistics (average time, average progress, etc.) to applying machine learning algorithms to make predictions and act proactively. In other words, as opposed to the descriptive analytic capabilities of LMSs, learning analytics should be predictive.

Although the identified potential of LA for improving learning support and outcomes is highly promising, the transfer of this potential into practice falls short of what was expected. This raises a question of how this transfer can be facilitated to ultimately benefit learners. Learning analytics is expected to empower teachers and students to understand and make better use of the abundance of data relevant to their learning. Through involvement in this process, teachers and students, and the entire institution, are given the opportunity to control their agenda to a greater extent and much more successfully, in a way that economic framing is complemented by care for learning. It is neither a simple nor a direct process. In order to achieve institutional change, it is not enough to focus only on data, it is necessary for data to be analysed and contextualized in ways that can initiate organizational change and development (Macfadyen & Dawson, 2012).

The application of LA process in educational institutions is often initiated by the demands and views of managers which is basically connected with the economic staffing of education. There is a gap in the perception of the role and value of LA between managers and teachers. And while managers usually give preference to the economic aspects of value, for teachers that value appears as having more information about their students. Satisfying the management demands, the LA environment of an education institution will certainly include metrics that measure the economic success of the learning process, but teachers should not be burdened with these metrics, but should try to understand the strengths and limitations of tools and techniques, and use this understanding for directing learning analytics to its basic purpose: improving teaching and learning.

Additional research is needed to improve the understanding of how higher education can benefit from the application of learning analytics, with particular focus on opportunities, barriers and challenges. This paper reviews the current state of art in learning analytics, which is fraught with many questions, issues, and disappointments regarding the observed effects. After outlining the most significant problems identified in the theory and practice of LA, we provided several practical guidelines and recommendations that could benefit HEIs that are starting or considering starting their LA endeavours.**SM**

References

Atkisson, M., & Wiley, D. (2011). Learning analytics as interpretive practice: applying Westerman to educational intervention. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 117-121). ACM. https://doi.org/10.1145/2090116.2090133

Bakharia, A., & Dawson, S. (2011). SNAPP: a bird's-eye view of temporal participant interaction. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 168-173). ACM. https://doi.org/10.1145/2090116.2090144

Broadbent, J., & Poon, W. (2015). Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1–15. https://doi.org/10.1016/j.iheduc.2015.04.007

Campbell, J., De Blois, P., & Oblinger, D. (2007). Academic Analytics: A New Tool for a New Era. *Educause Review, 42*(4), 40-57.

Clow, D. (2012). The learning analytics cycle: closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge,* 134-138. ACM. https://doi.org/10.1145/2330601.2330636

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology and Society, 11*(3), 224–238.

Ferguson, R. (2012). Learning analyitics: drivers, developments and challenges. *Internatonal Journal of Technology Enhanced Learning, 4*(5/6), 304-317. https://doi.org/10.1504/IJTEL.2012.051816

Ferguson, R., Clow, D., Griffiths, D. & Brasher, A. (2019). Moving forward with learning analytics: expert views, *Journal of Learning Analytics, 6*(3), 43-59. https://doi.org/10.18608/jla.2019.63.8

Francis, P. Broughan, C. Foster, C. & Wilson C. (2020). Thinking critically about learning analytics, student outcomes, and equity of attainment. *Assessment & Evaluation in Higher Education, 45*(6), 811-821. https://doi.org/10.1080/02602938.2019.1691975

Lee, LK. (2017). Rethinking the accessibility of online higher education: a historical overview. *The Internet and Higher Education, 33*, 15–23. https://doi.org/10.1016/j.iheduc.2017.01.001

Long, P., & Siemens, G. (2011). Penetrating the fog: analytics in learning and education. *Educause Review, 46*(5), 31-40.

Macfadyen, L., & Dawson. S. (2012). Numbers are not enough. why eLearning analytics failed to inform an institutional strategic plan. *Educational Technology and Society, 15*(3), 149–163.

Mah, D.-K. (2016). Learning analytics and digital badges: Potential impact on student retention in higher education. *Technology, Knowledge and Learning, 21*(3), 285–305. https://doi.org/10.1007/s10758-016-9286-8

Omedes, A., J. (2018). Learning Analytics 2018 – an updated perspective. IADLearning. Retrieved: November 18, 2021, from https://www.iadlearning.com/learning – analytics – 2018/

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior, 7*, 397–407. https://doi.org/10.1016/j.chb.2017.06.030

Sønderlund, A.L., Hughes, E. & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology, 5*(5), 2594-2618. https://doi.org/10.1111/bjet.12720

Strang, K. (2016). Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes? *Education and Information Technologies, 20*(4), 1–21

Suthers, D., & Chu, K. (2012). Multi-mediated community structure in a sociotechnical network. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 43-53). ACM. https://doi.org/10.1145/2330601.2330618

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior 89*, 98–110. https://doi.org/10.1016/j.chb.2018.07.02

✉ **Correspondence**

Veselin Pavlićević

University of Novi Sad, Faculty of Economics in Subotica, Subotica, Serbia
Segedinski put 9-11, 24000, Subotica, Serbia

E-mail: pavlicevic@ef.uns.ac.rs