Chapter 1
Learning Analytics in Higher Education—A Literature Review

Philipp Leitner, Mohammad Khalil and Martin Ebner

Abstract This chapter looks into examining research studies of the last five years and presents the state of the art of Learning Analytics (LA) in the Higher Education (HE) arena. Therefore, we used mixed-method analysis and searched through three popular libraries, including the Learning Analytics and Knowledge (LAK) conference, the SpringerLink, and the Web of Science (WOS) databases. We deeply examined a total of 101 papers during our study. Thereby, we are able to present an overview of the different techniques used by the studies and their associated projects. To gain insights into the trend direction of the different projects, we clustered the publications into their stakeholders. Finally, we tackled the limitations of those studies and discussed the most promising future lines and challenges. We believe the results of this review may assist universities to launch their own LA projects or improve existing ones.

Keywords Learning analytics · Higher education · Stakeholders · Literature review

P. Leitner (✉) · M. Khalil (✉) · M. Ebner
Educational Technology, Graz University of Technology,
Münzgrabenstraße 35A/I, 8010 Graz, Austria
e-mail: philipp.leitner@tugraz.at

M. Khalil
e-mail: mohammad.khalil@tugraz.at

M. Ebner
e-mail: martin.ebner@tugraz.at

A. Peña-Ayala (ed.), *Learning Analytics: Fundaments, Applications, and Trends*, Studies in Systems, Decision and Control 94,
DOI 10.1007/978-3-319-52977-6_1

Abbreviations

AA	Academic analytics
ACM	Association for computing machinery
EDM	Educational data mining
HE	Higher education
ITS	Intelligent tutoring system
LA	Learning analytics
LAK	Learning analytics and knowledge
LMS	Learning management system
MOOC	Massive open online course
NMC	New media consortium
PLE	Personal learning environment
RQ	Research question
SNA	Social network analysis
VLE	Virtual learning environment
WOS	Web of science

1.1 Introduction

The aim of LA is to evaluate user's behavior in the context of teaching and learning, further to analyze and interpret it to gain new insights and to provide the stakeholders with new models for improving teaching, learning, effective organization, and decision making (Siemens and Long 2011). A key fact is the return of the resulting knowledge to the teachers and students to optimize their teaching and learning behavior, to promote the development of skills in the area, and to better understand education as well as the connected fields, e.g. university business and marketing. Available resources can be used more efficiently to provide better support and individual care to develop potentials.

In the area of HE, LA has proven to be helpful to colleges and universities in strategic areas such as resource allocation, student success, and finance. These institutions are collecting more and more data than ever before, to maximize strategic outcomes. Based on key questions data is analyzed and predictions are made to gain insights and set actions. Many examples of successful analytics and frameworks use are available across a diverse range of institutions (Bichsel 2012). Ethical and legal issues of collecting and processing students' data are seen as barriers by the HE institutions in LA (Sclater 2014).

In this chapter, we present a literature review to evaluate the progress of LA in HE since its early beginning in 2011. We conducted the search with the three popular libraries: the LAK conference, the SpringerLink, and the WOS databases.

We then refined the returned results and settled on including 101 relevant publications. This chapter mainly contributes by analyzing them and lists the used

LA methods, limitations and stakeholders. It is expected that this study will be a guide for academicians who would like to improve existing LA projects or assist universities to launch their own.

The next section gives a short introduction on the topic of LA and describes LA in HE in detail. The subsequent sections are concerned with our research design, methodology and execution of the review. The outcomes of the research questions and the literature survey are presented in the third section. The penultimate section discusses the findings and shows the conclusion of our survey. A glance of future trends are presented in the last section.

1.2 A Profile of Learning Analytics and Learning Analytics in Higher Education

In this section we present a profile of LA in general and describe the analysis process. Further, we give emphasis to LA in HE, discuss challenges and identify the involved stakeholders.

1.2.1 Learning Analytics

Since its first mention in the New Media Consortium (NMC) Horizon Report 2012 (Johnson et al. 2012), LA has gained an increasing relevance. LA is 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" (Elias 2011). Another definition states "the use of intelligent data, learner-produced data, and analysis models to discover information and social connection, and to predict and advise on learning" (Siemens 2010).

The NMC Horizon Report 2013 identified LA as one of the most important trends in technology-enhanced learning and teaching (Johnson et al. 2013). Therefore, it is not surprising, that LA is the subject of many scientific papers. The research and improvement of LA involves doing the development, the use and integration of new processes and tools to improve the performance of teaching and learning of individual students and of teachers. LA focuses specifically on the process of learning (Siemens and Long 2011).

Due to its connections with digital teaching and learning, LA is an interdisciplinary research field with connections to the field of teaching and learning research, computer science and statistics (Johnson et al. 2013). The available data is collected, analyzed and the gained insights are used to understand the behavior of the students to provide them additional support (Gašević et al. 2015).

A key concern of LA is the gathering and analyzation of data as well as the setting of appropriate interventions to improve the learners learning experience (Greller

Fig. 1.1 The five steps of the analysis process

et al. 2014). These "actionable intelligence" from Educational Data Mining (EDM) is supporting the teaching and learning and provides ideas for customization, tutoring and intervention within the learning environment (Campbell et al. 2007).

According to Campbell and Oblinger (Campbell and Oblinger 2007), an analysis process has five steps, shown in Fig. 1.1.

Capturing, data is captured and collected in real-time from different sources like Virtual Learning Environments (VLE), Learning Management Systems (LMS), Personal Learning Environment (PLE), web portals, forums, chat or rooms, and combined with student information (Lauría et al. 2012; Tseng et al. 2016).

Reporting, the collected data is used to generate accurate models for identifying and measuring the student's progress. Often visualization is used in LA dashboards for a better understanding of the data (Muñoz-Merino et al. 2013; Leony et al. 2013).

Predicting, the data is used to identify predictors for student success, outcomes and for identifying at-risk students. Further, it is used for decision-making about courses and resource allocation which then is used by the decision-makers of the institutions (Akhtar et al. 2015; Lonn et al. 2012).

Acting, the information gained from the data analyzation process is used to set appropriate interventions in e.g. teaching or supporting students who are at risk of failure or dropping out (Freitas et al. 2015; Palmer 2013).

Refining, the gathered information is used in a cyclical process for continuous improvements of the used model in teaching and learning (Nam et al. 2014; Pistilli et al. 2014).

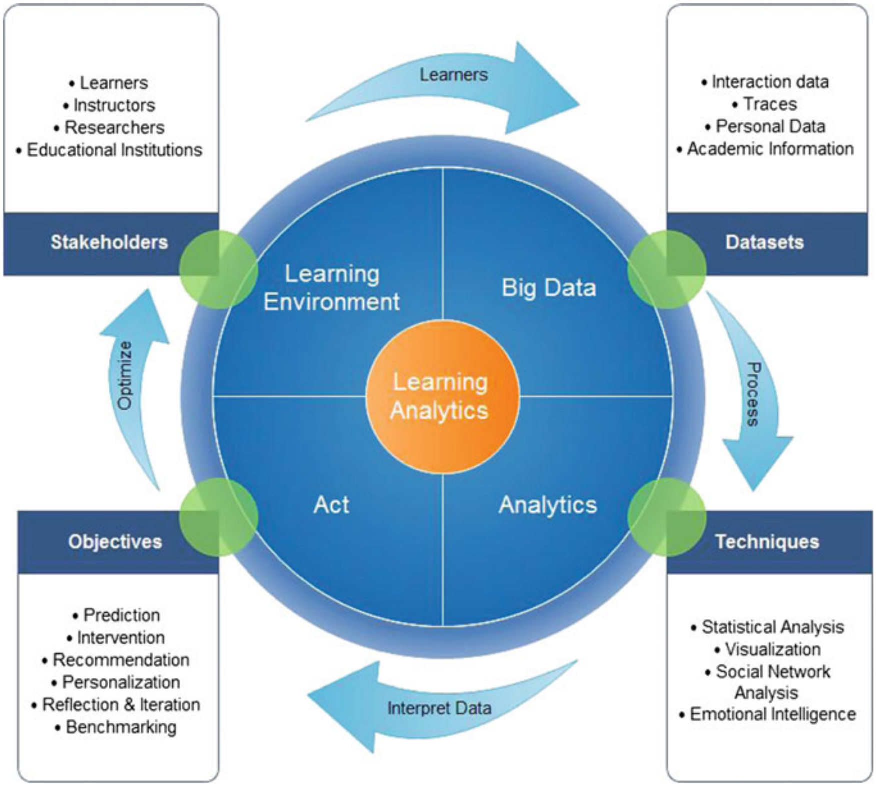


Fig. 1.2 Khalil and Ebner LA life cycle (Khalil and Ebner 2015)

Although research in the field of LA in recent years celebrates boom, LA is still in its infancy. Students, researchers and educational managers need to discuss ideas and opportunities on how to integrate these possibilities in their research and practice (Ferguson 2012).

In 2015, an LA approach which depicted a life cycle was introduced by Khalil and Ebner (2015), as shown in Fig. 1.2.

The cycle includes four main stages:

- Generation of data: this process starts from the learning environments where different stakeholders reside in MOOC, LMSs or any other VLEs.
- Data storage: learners leave a lot of traceable data behind them. Learners are not just consumers but also producers of data.
- Analysis: analytics methods seek to discover hidden patterns inside educational datasets. Analytics techniques are various. The authors defined them mainly into quantitative and qualitative analysis methods.

- Act: the analysis outcome should be interpreted to actions. In this stage, action is considered as prediction, intervention, recommendation, personalization and reflection.

At the end, the life cycle loop is closed by introducing the "optimization" process. Similar to Campbell and Oblinger (2007), they realized that there are similarities in phases between the available LA approaches in the literature. LA is an open loop of stages that should be closed at the end by optimizing learning environments and stakeholders (learners, tutors, decision makers…etc.).

1.2.2 Learning Analytics in Higher Education

HE looks forward to a future of uncertainty and change. In addition to the national and global as well as political and social changes, the competition on university level increases.

HE needs to increase financial and operational efficiency, expand local and global impact, establish new funding models during a changing economic climate and respond to the demands for greater accountability to ensure organizational success at all levels (van Barneveld et al. 2012). HE must overcome these external loads in an efficient and dynamic manner, but also understand the needs of the student body, who represents the contributor as well as the donor of this system (Shacklock 2016).

In addition to the strong competition, universities have to deal with the rapidly changing technologies that have arisen with the entry of the digital age. In the course of this, institutions collected enormous amounts of relevant data as a by-product. For instance, when students take an online course, use an Intelligent Tutoring System (ITS) (Arnold and Pistilli 2012; Bramucci and Gaston 2012; Fritz 2011; Santos et al. 2013) play educational games (Gibson and de Freitas 2016; Holman et al. 2013, 2015; Westera et al. 2013) or simply use an online learning platform (Casquero et al. 2014, 2016; Wu and Chen 2013; Ma et al. 2015; Santos et al. 2015; Softic et al. 2013).

In recent years, more universities use methods of LA in order to obtain findings on the academic progress of students, predict future behaviors and recognize potential problems in an early stage. Further, LA in the context of HE is an appropriate tool for reflecting the learning behavior of students and provide suitable assistance from teachers or tutors. This individual or group support offers new ways of teaching and provides a way to reflect on the learning behavior of the student.

Another motivation behind the use of LA in universities is to improve the inter-institutional cooperation, and the development of an agenda for the large community of students and teachers (Atif et al. 2013).

Table 1.1 Overview of the stakeholders (Romero and Ventura 2013)

Stakeholder	Objectives, benefits and perspectives
Learner	Support the learner with adaptive feedback, recommendations, response to his or her needs, for learning performance improvement
Educators	Understand students' learning process, reflect on teaching methods and performance, understand social, cognitive and behavioral aspects
Researchers	Use the right EDM technique which fits the problem, evaluation of learning effectiveness for different settings
Administrators	Evaluation of institutional resources and their educational offer

On an international level, the recruitment, management and retention of students have become as high level priorities for decision makers in institutions of HE. Especially improving the student retention starts and the understanding of the reason behind and/or prediction of the attrition has come in the focus of attention due to the financial losses, lower graduation rates, and inferior school reputation in the eyes of all stakeholders (Delen 2010; Palmer 2013).

Despite that LA focuses strongly on the learning process, the results still in the beneficial for all stakeholders. Romero and Ventura (2013) divided those involved stakeholders based on their objectives, benefits and perspectives in the four groups shown in Table 1.1.

1.3 Research Design, Methodology and Execution

This research aims at the elicitation of an overview on the advancement of the LA field in HE since it emerged in 2011. The proposed Research Questions (RQ) to answer are:

- **RQ1**: What are the research strands of the LA field in HE (between January 2011 and February 2016)?
- **RQ2**: What kind of limitations do the research papers and articles mention?
- **RQ3**: Who are the stakeholders and how could they be categorized?
- **RQ4**: What methods do they use in their papers?

1.3.1 Literature Review Procedure

In accordance to this objective, we performed a literature review following the procedure of Machi and McEvoy (2009). Figure 1.3 displays the six steps for a literature review used in this process.

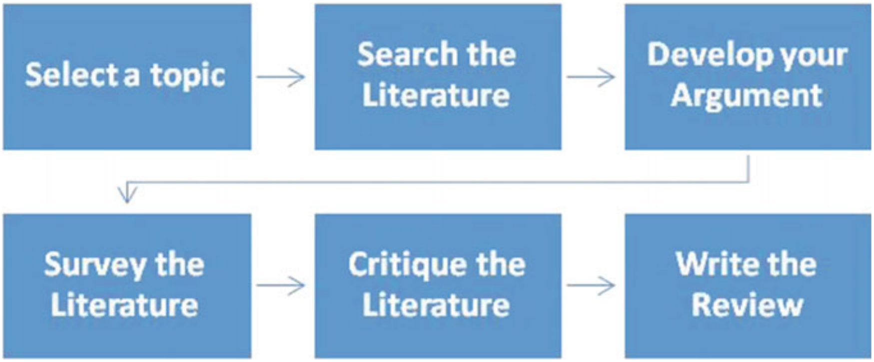


Fig. 1.3 The literature review: six steps to success (Machi and McEvoy 2009)

After we selected our topic, we identified data sources based on their relevance in the computing domain:

- The papers of the LAK conference published in the Association for Computing Machinery (ACM) Digital Library,
- The SpringerLink, and
- The Thomson Reuters WOS database.

and the following search parameters:

In the LAK papers, we didn't need to search for the "Learning Analytics" term because the whole conference covers the LA discipline. We searched the title, the abstract and the author keywords for "Higher Education" and/or "University".

In the SpringerLink database, we searched for the "Learning Analytics" term in conjunction with either "Higher Education" or "University" ("Learning Analytics AND (Higher Education OR University)").

In the WOS database, we searched for the topic "Learning Analytics" in conjunction with either "Higher Education" or "University" and in the research domain "science technology".

The defined inclusion criteria of the fetched papers from the libraries were set to be: (a) written in English, and (b) published between 2011 till the February 2016. We superficially assessed the quality of the reported studies, considering only articles that provided substantial information for LA in HE. Therefore, we excluded articles that did not meet the outlined inclusion principles.

The literature survey was conducted in February and March 2016. In the initial search, we found a total of 135 publications (LAK: 65, SpringerLink: 37, WOS: 33). During the first stage, the search results were analyzed based on their titles, author keywords and abstracts. After this stage, 101 papers remain for the literature survey. We fully read each publication and actively searched for their research questions, techniques, stakeholders, and limitations. Regular meetings between the authors were set on a weekly basis to discuss the results. Additionally, we added to our spreadsheet the Google Scholar[1] citation count as a measurement of article's impact.

In order to present our findings, we analyze each of the research questions separately. This section presents our findings.

1.3.2 Response to Research Question 1

In order to answer the RQ1, which corresponds to "What are the research strands of the LA field in HE (between January 2011 and February 2016)?", we tried to extract the main topics from the research questions of the publications.

We identified that many of the publications do not outline their research questions clearly. Many of the examined publications described use cases. This concerns in particular the older publications of 2011 and 2012, and is probably resulting from the young age of the scientific field of LA.

As a result, we did a brief text analysis on the fetched abstracts in order to examine the robust trends in the prominent field of LA and HE. We have collected all the article abstracts, processed them through the R software, and then refined the resulted corpus. In the final stages, we demonstrated the keywords and chose the Word cloud as a representation tool of the terms as shown in Fig. 1.4. The figure was graphically generated using one of the R library packages called "wordcloud".[2]

In order to ease reading the cloud, we adopted four levels of representation depicted in four colors. The obtained list of words that have been used were classified into singular phrases, bi-grams, tri-grams and quad-grams. The most cited singular words were "academic", "performance", "behavior" and "MOOCs". "Learning environment", "case study" and "online learning" were the most repeated bi-grams. The highest tri-grams used in the abstracts were "learning management systems", "Higher Education institutions" and "social network analysis". While quad-grams were only limited to "massive open online courses" which were merged at the final filtering stage with the "MOOCs" term.

The word cloud shows a glance about the general topics when LA is ascribed with HE. LA researchers focused on utilizing its techniques towards enhancing performance and students' behaviors. The popular adopted educational environment was Massive Open Online Course (MOOC) platforms. Furthermore, LA was also used to perform practices of interventions, observing dropout, videos, dashboards and engagement.

In Fig. 1.5 the collected articles are from the library data sources. Results show an obvious increase in the number of publications since 2011. For instance, there were 32 papers in 2015, incremented from 26 articles in 2014 and 17 articles in 2013. However, there were 5 articles only in 2011 and 12 articles in 2012. Because

[1]Online: http://scholar.google.com.

[2]Online: https://cran.r-project.org/web/packages/wordcloud/index.html.

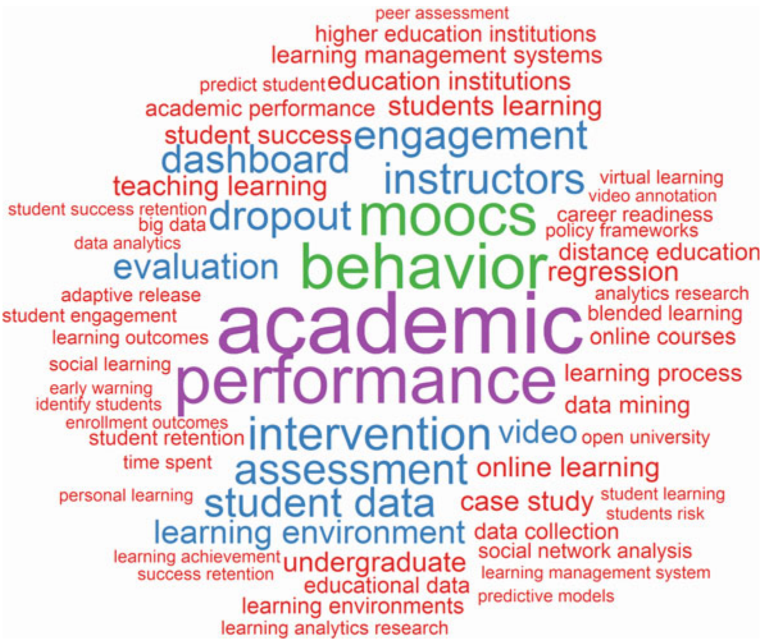
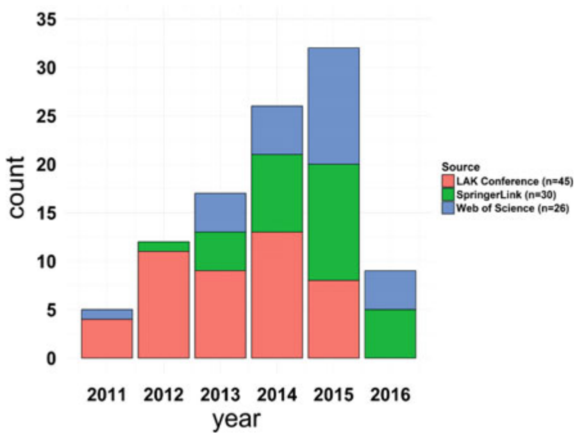


Fig. 1.4 Word cloud of the prominent terms from the abstracts

Fig. 1.5 Collected articles distributed by source and year



February 2016 was the date of collecting the publications in this study, the 2016 year was not indexed with many papers. On the other hand, the figure shows the apparent involvement of the journal articles from the SpringerLink and WOS libraries from 2013.

We cross-referenced the relevant publications with Google Scholar to derive their citation impact. Table 1.2 shows the 10 most cited publications.

Table 1.2 Citation impact of the publications

Paper title	Year of publication	No. of Google citations (Feb. 2016)
Course signal at Purdue: using learning analytics to increase student success (Arnold and Pistilli 2012)	2012	164
Social learning analytics: five approaches (Ferguson and Shum 2012)	2012	94
Classroom walls that talk: using online course activity data of successful students to raise self-awareness of underperforming peers (Fritz 2011)	2011	52
Goal-oriented visualizations of activity tracking: a case study with engineering students (Santos et al. 2012)	2012	46
Where is research on massive open online courses headed? A data analysis of the MOOC research initiative (Gasevic et al. 2014)	2014	46
Course correction: using analytics to predict course success (Barber and Sharkey 2012)	2012	36
Improving retention: predicting at-risk students by analyzing clicking behavior in a virtual learning environment (Wolff et al. 2013)	2013	34
Learning designs and learning analytics (Lockyer and Dawson 2011)	2011	33
The pulse of learning analytics understandings and expectations from the stakeholders (Drachsler and Greller 2012)	2012	30
Inferring higher level learning information from low level data for the Khan Academy platform (Muñoz-Merino et al. 2013)	2013	28

1.3.3 Response to Response to Research Question 2

We identified for RQ2, which corresponds to "What kind of limitations do the research papers and articles mention?", three different limitations, either clearly mentioned in articles or being tacitly within the context.

Limitations through time, some of the publications stated that continuous work is needed (Elbadrawy et al. 2015; Ifenthaler and Widanapathirana 2014; Koulocheri and Xenos 2013; Lonn et al. 2012; Palavitsinis et al. 2011; Sharkey 2011). Either a longitudinal study would be necessary to prove hypotheses or because of the shortage of the project (Fritz 2011; Nam et al. 2014; Ramírez-Correa and Fuentes-Vega 2015).

Limitations through the size, other publications talked about the need for more detailed data (Barber and Sharkey 2012; Best and MacGregor 2015; Rogers et al. 2014), the small group sizes (Junco and Clem 2015; Jo et al. 2015; Martin and

Whitmer 2016; Strang 2016), the unsure scalability, possible problems in wider context and the problem of the generalization of the approach or method (Prinsloo et al. 2015; Yasmin 2013).

Limitations through the culture, many of the publications mention that their approach might only work in their educational culture and is not applicable somewhere else (Arnold et al. 2014; Drachsler and Greller 2012; Grau-Valldosera and Minguillón 2014; Kung-Keat and Ng 2016). Additionally, the ethics differ strongly around the world, so cooperation projects between different universities in different countries needs different moderation as well as the use of data could be ethically questionable (Abdelnour-Nocera et al. 2015; Ferguson and Shum 2012; Lonn et al. 2013; Park et al. 2016).

Furthermore, ethical discussions about data ownership and privacy have recently arisen. Slade and Prinsloo (2013) pointed out that LA touches various research areas and therefore overlaps with ethical perspectives in areas of data ownership and privacy. Questions about who should own the collected and analyzed data were highly debated. As a result, the authors classified the overlapping categories in three parts:

- The location and interpretation of data,
- Informed consent, privacy and the de-identification of data, and
- The management, classification and storage of data.

These three elements generate an imbalance of power between the stakeholders which they addressed by proposing a list of 6 grounding principles and considerations: LA as moral practice, students as agents, student identity and performance are temporal dynamic constructs, student success is a complex and multidimensional phenomenon, transparency, HE cannot afford to not use data (Slade and Prinsloo 2013).

1.3.4 Response to Response to Research Question 3

In order to answer the RQ3, which corresponds to "Who are the stakeholders and how could they be categorized?", we determined the stakeholders from the publications and categorized them into three types. As a basis, we took the four stakeholders as mentioned in Sect. 1.2.2 and introduced in (Machi and McEvoy 2009). We merged the Researchers and Administrators from the original classification into one distinct group. Therefore, the institutional perspective [Academic Analytics (AA)] is separated from the learners' and teachers' one (LA).

Figure 1.6 depicts the defined LA stakeholders as a Venn-Diagram. The figure shows that there had been more research conducted concerning the Researchers/Administrators with overall 65 publications and 40 of them only concerning themselves, than in the field of Learners with a total of 53 publications and 21 single mentions. Also, it seems that Teachers are only a "side-product" of this field with only 20 mentions and only 7 dedicated to them alone.

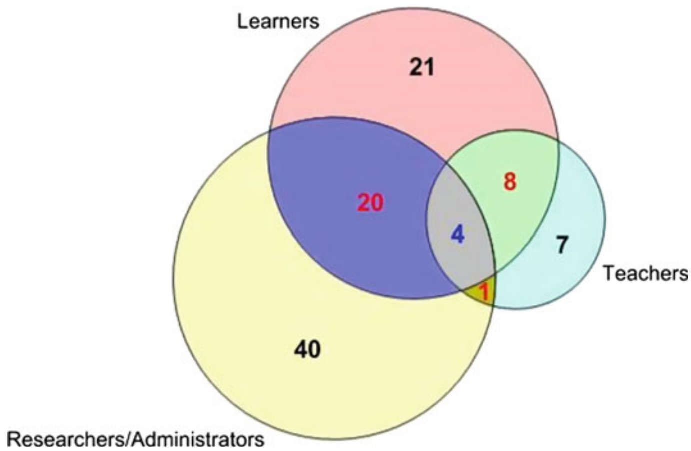


Fig. 1.6 Venn-diagram of stakeholders in the publications

Most of the combined articles addressed Researchers/Administrators together with Learners (20 publications). Only 8 articles can be found with an overlap between Learners and Teachers, which should be one of the most researched and discussed combinations within LA in HE. Nearly no work has been done by combining Researchers/Administrators with Teachers (in 1 publications) and only 4 paper combined all 3 stakeholders. This lack of research will be a matter of debate in the discussion section.

1.3.5 Response to Response to Research Question 4

By analyzing the selected studies to answer RQ4, which corresponds to "What techniques do they use in their papers?", we identified the techniques used in LA and HE publications. We took into account the methods presented by Romero and Ventura (2013), Khalil and Ebner (2016) and Linan and Perez (2015). We propose an overview of the used techniques of the different articles in Table 1.3.

The results of Fig. 1.7 show, that the research is focused mainly on prediction with a total of 36 citations. Outlier detection for pointing out at-risk or dropping out students with a citation count of 29. Distillation of data for human judgment in form of a visualization with a citation count of 33 than in all other parts including rarely used techniques like gamification or machine learning with a total amount of 102 counts.

Table 1.3 Overview of the used LA techniques of this study

Techniques	Key applications	Examples
Prediction	Predicting student performance and detecting student behaviors	AbuKhousa and Atif (2016), Cambruzzi et al. (2015), Harrison et al. (2015)
Clustering	Grouping similar materials or students based on their learning and interaction patterns	Aguiar et al. (2014), Asif et al. (2015), Scheffel et al. (2012)
Outlier detection	Detection of students with difficulties or irregular learning processes	Grau-Valldosera and Minguillón (2011), Manso-Vázquez and Llamas-Nistal (2015), Sinclari and Kalvala (2015)
Relationship mining	Identifying relationships in learner behavior patterns and diagnosing student difficulties	Kim et al. (2016), Pardo et al. (2015), Piety et al. (2014)
Social network analysis	Interpretation of the structure and relations in collaborative activities and interactions with communication tools	Hecking et al. (2014), Tervakari et al. (2013), Vozniuk et al. (2014)
Process mining	Reflecting student behavior in terms of its examination traces, consisting of a sequence of course, grade and timestamp	Menchaca et al. (2015), Vahdat et al. (2015), Wise (2014)
Text mining	Analyzing the contents of forums, chats, web pages and documents	Gasevic et al. (2014), Lotsari et al. (2014), Prinsloo et al. (2012)
Distillation of data for human judgment	Helping instructors to visualize and analyze the ongoing activities of the students and the use of information	Aguilar et al. (2014), Grann and Bushway (2014), Swenson (2014)
Discovery with models	Identification of relationships among student behaviors and characteristics or contextual variables. Integration of psychometric modelling frameworks into machine-learning models	Gibson et al. (2014), Kovanović et al. (2015), Lockyer and Dawson (2011)
Gamification	Include possibilities for playful learning to maintain motivation; e.g. integration of achievements, experience points or badges as indicators of success	Holman et al. (2013), Øhrstrøm et al. (2013), Westera et al. (2013)
Machine learning	Find hidden insights in data automatically (based on models who are exposed to new data and adapt itself independently)	Corrigan et al. (2015), McKay et al. (2012), Nespereira et al. (2016)
Statistic	Analysis and interpretation of quantitative data for decision making	Clow (2014), Khousa and Atif (2014), Simsek et al. (2015)

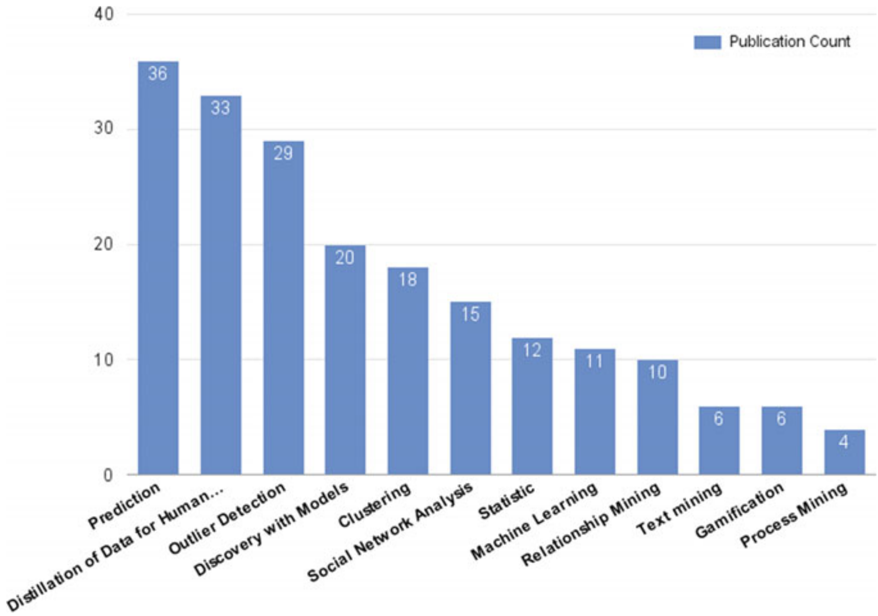


Fig. 1.7 The publication count of the used LA techniques

1.4 Discussion and Conclusion

In this chapter, we examined hundreds of pages to introduce a remarkable literature review of the LA field in the HE domain. We presented a state-of-the-art study of both domains based on analyzing articles from three major library references: the LAK conference, SpringerLink and WOS. The total number of relevant publications was equal to 101 articles in a period between 2011 and 2016.

In this literature review study, we followed the procedure of Machi and McEvoy (2009) in which we selected the topic, searched the literature to get the answers to the research questions, surveyed and critiqued the literature and finally introduced our review. Using this big dataset, we identified the research strands of the relevant publications. Most of the publications described use cases rather than comprehensive research—especially the prior publications, which is comprehensible because at the time, the universities had to figure out how to handle and harness the abilities offered by LA for their benefit.

To make a better holistic overview on the advancement of LA field in HE, we proposed four main RQs. These questions were related to the research strands of LA in HE, limitations, stakeholders and what techniques were used by LA experts in the HE domain, respectively.

The first research question was answered by generating a word cloud of a final corpus which was formed from all abstracts of the included papers. Results revealed that the usage of MOOCs, enhancing learning performance, students behavior, and benchmarking learning environments were strongly researched by LA experts in the domain of HE.

The paper with the title "Course signals at Purdue: using learning analytics to increase student success" by Arnold and Pistilli (2012), was the most cited article of our inclusion, which focused on a tool of prediction. Also, we identified that there was a clear increment of publications since 2011 till 2015, Further it was shown the apparent involvement of the journal articles from the SpringerLink and WOS libraries in 2013 and 2015 over the LAK conference publications.

The second research questions showed that limitations were mainly concerning the needed time to prepare data or getting the results, the size of the available dataset and examined group and ethical reasons. While the discussions of privacy and ownership have arisen dramatically after 2012, we found that the ethical constraints drive the limitations to the greatest extent of this literature review study similar to the arguments in Khalil and Ebner (2015, 2016b).

The analysis shows that there was clamor regarding who are the main stakeholders of LA and HE. As the leading stakeholders of LA should be learners and students (Khalil and Ebner 2015), we found that researchers play a major role of the loop between HE and LA. Figure 1.6 demonstrated the high use of researchers and administrators in carrying out decisions. The direct overlap between learners and teachers was not evidently identified in our study.

At the final stage, we tried to elaborate what were the most used techniques of LA in HE. This research question was answered based on solid articles that discussed the LA techniques. The scanning showed that prediction, distilling of data for human judgment, and outlier detection were the most used methods in the HE domain. General data mining methodologies from text mining to Social Network Analysis (SNA) were identified with high usage in the analyzed publications. On the other hand, we noticed that there are new techniques that seem to be used more frequently in the past two years such as serious gaming, which belongs to the gamification techniques.

1.5 Future Trends

In this section we tackle the future development in the field of LA in HE, which can be divided into short-term (1–2 years) and long term (3–5 years) trends.

1.5.1 Short-Term Trends

Over the next 1 to 2 years, universities must adjust to the social and economic factors, which postulated the change in the capabilities of the students (Johnson et al. 2016). The tuning of the areas analysis, consultation, examination of individual learning outcomes and the visualization of continuously-available, aggregated information in dashboards are gaining more and more importance. Students expect real-time feedback during learning with critical self-reflection on the learning progress and learning goal which strengthens their expertise in self-organization. If adequate quantities of data from students are available, they can be carried out for subsequently, predictive analytics (Johnson et al. 2016).

1.5.2 Long-Term Trends

The relevance of LA in HE will mint even more over the next 3–5 years. This trend is promoted by the strong interest of students for individual evaluations and care. To serve this market, dashboards and analysis applications that specifically address the needs of each customer will develop stronger. This approach offers many advantages: Accessing your own data in an appropriate form allows better self-reflection and a healthy rivalry among the fellow students. The teachers can survey a large amount of students and precisely recognize those who need their help. University and college dropouts can be better detected by appropriate analyzing and with targeted interventions they remain in the university system (Shacklock 2016).

To master the associated problems, the LA market will have to change. Currently, many different systems and analytical approaches are used. The fragmentation of the market will grow even further in the future, which makes the interuniversity comparison very difficult or even impossible. Therefore, the creation of standards is essential (Shacklock 2016). Furthermore, a change in the type of analysis is foreseeable. Most current and past data have been used to measure the success of students. Today, advances in predictive analytics (predictive analysis) are important. By using the analysis of existing data sets of many students, predictive models can be developed and warn thus students who are at risk not to meet their learning success (Shacklock 2016).

Acknowledgements This research project is co-funded by the European Commission Erasmus+ program, in the context of the project 562167-EPP-1-2015-1-BE-EPPKA3-PI-FORWARD.

References

Abdelnour-Nocera J, Oussena S, Burns C (2015) Human work interaction design of the smart university. In: Human work interaction design. Work analysis and interaction design methods for pervasive and smart workplaces. Springer International Publishing, pp 127–140

AbuKhousa E, Atif Y (2016) Virtual social spaces for practice and experience sharing. In: State-of-the-Art and Future Directions of Smart Learning. Springer, Singapore, pp 409–414

Aguiar E, Chawla NV, Brockman J, Ambrose GA, Goodrich V (2014) Engagement vs performance: using electronic portfolios to predict first semester engineering student retention. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 103–112

Aguilar S, Lonn S, Teasley SD (2014) Perceptions and use of an early warning system during a higher education transition program. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 113–117

Akhtar S, Warburton S, Xu W (2015) The use of an online learning and teaching system for monitoring computer aided design student participation and predicting student success. Int J Technol Des Edu, pp 1–20

Arnold KE, Pistilli MD (2012) Course signals at Purdue: using learning analytics to increase student success. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 267–270

Arnold KE, Lonn S, Pistilli MD (2014) An exercise in institutional reflection: the learning analytics readiness instrument (LARI). In: Proceedings of the fourth international conference on learning penetrating the black box of time-on-task estimation and knowledge. ACM, pp 163–167

Asif R, Merceron A, Pathan MK (2015) Investigating performance of students: a longitudinal study. In: Proceedings of the fifth international conference on learning analytics and knowledge. ACM, pp 108–112

Atif A, Richards D, BilginA, Marrone M (2013) Learning analytics in higher education: a summary of tools and approaches. In: 30th Australasian Society for computers in learning in tertiary education conference, Sydney

Barber R, Sharkey M (2012) Course correction: using analytics to predict course success. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 259–262

Best M, MacGregor D (2015) Transitioning design and technology education from physical classrooms to virtual spaces: implications for pre-service teacher education. Int J Technol Des Edu, pp 1–13

Bichsel J (2012) Analytics in higher education: benefits, barriers, progress, and recommendations. EDUCAUSE Center for Applied Research

Bramucci R, Gaston J (2012) Sherpa: increasing student success with a recommendation engine. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 82–83

Cambruzzi WL, Rigo SJ, Barbosa JL (2015) Dropout prediction and reduction in distance education courses with the learning analytics multitrail approach. J UCS 21(1):23–47

Campbell JP, Oblinger DG (2007) Academic analytics, EDUCAUSE white paper. Retrieved 10 Feb 2016 from https://net.educause.edu/ir/library/pdf/PUB6101.pdf

Campbell JP, DeBlois PB, Oblinger DG (2007) Academic analytics: a new tool for a new era. EDUCAUSE Rev 42(4):40–57

Casquero O, Ovelar R, Romo J, Benito M (2014) Personal learningenvironments, highereducation and learninganalytics: a study of theeffects of servicemultiplexityonundergraduatestudents' personal networks/Entornos de aprendizaje personales, educación superior y analítica del aprendizaje: un estudio sobre los efectos de la multiplicidad de servicios en las redes personales de estudiantes universitarios. Cultura y Educación 26(4):696–738

Casquero O, Ovelar R, Romo J, Benito M, Alberdi M (2016) Students' personal networks in virtual and personal learning environments: a case study in higher education using learning analytics approach. Interact Learning Environ 24(1):49–67

Clow D (2014) Data wranglers: human interpreters to help close the feedback loop. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 49–53

Corrigan O, Smeaton AF, Glynn M, Smyth S (2015) Using educational analytics to improve test performance. In: Design for teaching and learning in a networked world. Springer International Publishing, pp 42–55

Delen D (2010) A comparative analysis of machine learning techniques for student retention management. Decis Support Syst 49(4):498–506
Drachsler H, Greller W (2012) The pulse of learning analytics understandings and expectations from the stakeholders. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 120–129
Elbadrawy A, Studham RS, Karypis G (2015) Collaborative multi-regression models for predicting students' performance in course activities. In: Proceedings of the fifth international conference on learning analytics and knowledge. ACM, pp 103–107
Elias T (2011) Learning analytics: definitions, processes and potential
Ferguson R (2012) Learning analytics: drivers, developments and challenges. Int J Technol Enhanced Learning 4(5/6):304–317
Ferguson R, Shum SB (2012) Social learning analytics: five approaches. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 23–33
Freitas S, Gibson D, Du Plessis C, Halloran P, Williams E, Ambrose M, Dunwell I, Arnab S (2015) Foundations of dynamic learning analytics: using university student data to increase retention. Br J Educational Technol 46(6):1175–1188
Fritz J (2011) Classroom walls that talk: using online course activity data of successful students to raise self-awareness of underperforming peers. Internet Higher Edu 14(2):89–97
Gasevic D, Kovanovic V, Joksimovic S, Siemens G (2014) Where is research on massive open online courses headed? A data analysis of the MOOC research initiative. Int Rev Res Open Distrib Learning, 15(5)
Gašević D, Dawson S, Siemens G (2015) Let's not forget: learning analytics are about learning. TechTrends 59(1):64–71
Gibson D, de Freitas S (2016) Exploratory analysis in learning analytics. Technol Knowl Learning 21(1):5–19
Gibson A, Kitto K, Willis J (2014) A cognitive processing framework for learning analytics. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 212–216
Grann J, Bushway D (2014) Competency map: visualizing student learning to promote student success. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 168–172
Grau-Valldosera J, Minguillón J (2011) Redefining dropping out in online higher education: a case study from the UOC. In: Proceedings of the 1st international conference on learning analytics and knowledge. ACM, pp 75–80
Grau-Valldosera J, Minguillón J (2014) Rethinking dropout in online higher education: The case of the UniversitatOberta de Catalunya. Int Rev Res Open Distrib Learning, 15(1)
Greller W, Ebner M, Schön M (2014) Learning analytics: from theory to practice–data support for learning and teaching. In: Computer assisted assessment. Research into e-assessment. Springer International Publishing, pp 79–87
Harrison S, Villano R, Lynch G, Chen G (2015) Likelihood analysis of student enrollment outcomes using learning environment variables: a case study approach. In: Proceedings of the fifth international conference on learning analytics and knowledge. ACM, pp 141–145
Hecking T, Ziebarth S, Hoppe HU (2014) Analysis of dynamic resource access patterns in a blended learning course. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 173–182
Holman C, Aguilar S, Fishman B (2013) GradeCraft: what can we learn from a game-inspired learning management system? In: Proceedings of the third international conference on learning analytics and knowledge. ACM, pp 260–264
Holman C, Aguilar SJ, Levick A, Stern J, Plummer B, Fishman B (2015) Planning for success: how students use a grade prediction tool to win their classes. In: Proceedings of the fifth international conference on learning analytics and knowledge. ACM, pp 260–264
Ifenthaler D, Widanapathirana C (2014) Development and validation of a learning analytics framework: two case studies using support vector machines. Technol Knowl Learning 19(1–2):221–240

Jo IH, Yu T, Lee H, Kim Y (2015) Relations between student online learning behavior and academic achievement in higher education: a learning analytics approach. In: Emerging issues in smart learning. Springer, Berlin, pp 275–287

Johnson L, Adams S, Cummins M (2012) The NMC horizon report: 2012 higher education edition. The New Media Consortium, Austin

Johnson L, Adams Becker S, Cummins M, Freeman A, Ifenthaler D, Vardaxis N (2013) Technology outlook for Australian tertiary education 2013–2018: an NMC horizon project regional analysis. New Media Consortium

Johnson L, Adams S, Cummins M, Estrada V, Freeman A, Hall C (2016) NMC horizon report: 2016 higher education edition. The New Media Consortium, Austin. http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf

Junco R, Clem C (2015) Predicting course outcomes with digital textbook usage data. Internet High Edu 27:54–63

Khalil M, Ebner M (2015) Learning analytics: principles and constraints. In: Proceedings of world conference on educational multimedia, hypermedia and telecommunications, pp 1326–1336

Khalil M, Ebner M (2016a) What is learning analytics about? A survey of different methods used in 2013–2015. In: Proceedings of smart learning conference, Dubai, UAE, 7–9 Mar. HBMSU Publishing House, Dubai, pp 294–304

Khalil M, Ebner M (2016b) De-identification in learning analytics. J Learning Anal 3(1), pp 129–138 http://dx.doi.org/10.18608/jla.2016.31.8

Khousa EA, Atif Y (2014) A learning analytics approach to career readiness development in higher education. In: International conference on web-based learning. Springer International Publishing, pp 133–141

Kim J, Jo IH, Park Y (2016) Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. Asia Pac Edu Rev 17(1):13–24

Koulocheri E, Xenos M (2013) Considering formal assessment in learning analytics within a PLE: the HOU2LEARN case. In: Proceedings of the third international conference on learning analytics and knowledge. ACM, pp 28–32

Kovanović V, Gašević D, Dawson S, Joksimović S, Baker RS, Hatala M (2015) Penetrating the black box of time-on-task estimation. In: Proceedings of the fifth international conference on learning analytics and knowledge. ACM, pp 184–193

Kung-Keat T, Ng J (2016) Confused, bored, excited? An emotion based approach to the design of online learning systems. In: 7th International conference on university learning and teaching (InCULT 2014) proceedings. Springer, Singapore, pp 221–233

Lauría EJ, Baron JD, Devireddy M, Sundararaju V, Jayaprakash SM (2012) Mining academic data to improve college student retention: an open source perspective. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 139–142

Leony D, Muñoz-Merino PJ, Pardo A, Kloos CD (2013) Provision of awareness of learners' emotions through visualizations in a computer interaction-based environment. Expert Syst Appl 40(13):5093–5100

Liñán LC, Pérez ÁAJ (2015) Educational data mining and learning analytics: differences, similarities, and time evolution. Revista de Universidad y SociedaddelConocimiento 12(3): 98–112

Lockyer L, Dawson S (2011) Learning designs and learning analytics. In: Proceedings of the 1st international conference on learning analytics and knowledge. ACM, pp 153–156

Lonn S, Krumm AE, Waddington RJ, Teasley SD (2012) Bridging the gap from knowledge to action: Putting analytics in the hands of academic advisors. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 184–18

Lonn S, Aguilar S, Teasley SD (2013) Issues, challenges, and lessons learned when scaling up a learning analytics intervention. In: Proceedings of the third international conference on learning analytics and knowledge. ACM, pp 235–239

Lotsari E, Verykios VS, Panagiotakopoulos C, Kalles D (2014) A learning analytics methodology for student profiling. In: Hellenic conference on artificial intelligence. Springer International Publishing, pp 300–312

Ma J, Han X, Yang J, Cheng J (2015) Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: the role of the instructor. Internet High Edu 24:26–34
Machi LA, McEvoy BT (2009) The literature review: six steps to success. Corwin Sage, Thousand Oaks
Manso-Vázquez M, Llamas-Nistal M (2015) A monitoring system to ease self-regulated learning processes. IEEE RevistaIberoamericana de TecnologiasdelAprendizaje 10(2):52–59
Martin F, Whitmer JC (2016) Applying learning analytics to investigate timed release in online learning. Technol Knowl Learning 21(1):59–74
McKay T, Miller K, Tritz J (2012) What to do with actionable intelligence: E 2 coach as an intervention engine. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 88–91
Menchaca I, Guenaga M, Solabarrieta J (2015) Project-based learning: methodology and assessment learning technologies and assessment criteria. In: Design for teaching and learning in a networked world. Springer International Publishing, pp 601–604
Muñoz-Merino PJ, Valiente JAR, Kloos CD (2013) Inferring higher level learning information from low level data for the Khan Academy platform. In: Proceedings of the third international conference on learning analytics and knowledge. ACM, pp 112–116
Nam S, Lonn S, Brown T, Davis CS, Koch D (2014) Customized course advising: investigating engineering student success with incoming profiles and patterns of concurrent course enrollment. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 16–25
Nespereira CG, Elhariri E, El-Bendary N, Vilas AF, Redondo RPD (2016) Machine learning based classification approach for predicting students performance in blended learning. In: The 1st International conference on advanced intelligent system and informatics (AISI2015), 28–30 Nov 2015, BeniSuef, Egypt. Springer International Publishing, pp 47–56
Øhrstrøm P, Sandborg-Petersen U, Thorvaldsen S, Ploug T (2013) Teaching logic through web-based and gamified quizzing of formal arguments. European conference on technology enhanced learning. Springer, Berlin, pp 410–423
Palavitsinis N, Protonotarios V, Manouselis N (2011) Applying analytics for a learning portal: the organic. Edunet case study. In: Proceedings of the 1st international conference on learning analytics and knowledge. ACM, pp 140–146
Palmer S (2013) Modelling engineering student academic performance using academic analytics. Int J Eng Educ 29(1):132–138
Pardo A, Mirriahi N, Dawson S, Zhao Y, Zhao A, Gašević D (2015) Identifying learning strategies associated with active use of video annotation software. In: Proceedings of the fifth international conference on learning analytics and knowledge. ACM, pp 255–259
Park Y, Yu JH, Jo IH (2016) Clustering blended learning courses by online behavior data: a case study in a Korean higher education institute. Internet High Educ 29:1–11
Piety PJ, Hickey DT, Bishop MJ (2014) Educational data sciences: framing emergent practices for analytics of learning, organizations, and systems. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 193–202
Pistilli MD, Willis III JE, Campbell JP (2014) Analytics through an institutional lens: definition, theory, design, and impact. In: Learning analytics. Springer New York, pp 79–102
Prinsloo P, Slade S, Galpin F (2012) Learning analytics: challenges, paradoxes and opportunities for mega open distance learning institutions. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 130–133
Prinsloo P, Archer E, Barnes G, Chetty Y, Van Zyl D (2015) Big (ger) data as better data in open distance learning. Int Rev Res Open Distrib Learning, 16(1)
Ramírez-Correa P, Fuentes-Vega C (2015) Factors that affect the formation of networks for collaborative learning: an empirical study conducted at a Chilean university/Factores que afectanla formación de redes para el aprendizajecolaborativo: unestudioempíricoconducidoenunauniversidadchilena. Ingeniare: RevistaChilena de Ingenieria, 23(3), 341

Rogers T, Colvin C, Chiera B (2014) Modest analytics: using the index method to identify students at risk of failure. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 118–122

Romero C, Ventura S (2013) Data mining in education. Wiley Interdiscip Rev Data Min Knowl Discovery 3(1):12–27

Santos JL, Govaerts S, Verbert K, Duval E (2012) Goal-oriented visualizations of activity tracking: a case study with engineering students. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, pp 143–152

Santos JL, Verbert K, Govaerts S, Duval E (2013) Addressing learner issues with StepUp!: an evaluation. In: Proceedings of the third international conference on learning analytics and knowledge. ACM, pp 14–22

Santos JL, Verbert K, Klerkx J, Duval E, Charleer S, Ternier S (2015) Tracking data in open learning environments. J Univ Comput Sci 21(7):976–996

Scheffel M, Niemann K, Leony D, Pardo A, Schmitz HC, Wolpers M, Kloos CD (2012) Key action extraction for learning analytics. European conference on technology enhanced learning. Springer, Berlin, pp 320–333

Sclater N (2014) Code of practice "essential" for learning analytics. http://analytics.jiscinvolve.org/wp/2014/09/18/code-of-practice-essential-for-learning-analytics/

Shacklock X (2016) From bricks to clicks: the potential of data and analytics in higher education. The Higher Education Commission's (HEC) report

Sharkey M (2011) Academic analytics landscape at the University of Phoenix. In: Proceedings of the 1st international conference on learning analytics and knowledge. ACM, pp 122–126

Siemens G (2010) What are learning analytics. Retrieved 10 Feb 2016 from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Siemens G, Long P (2011) Penetrating the fog: analytics in learning and education. EDUCAUSE Rev 46(5):30–40

Simsek D, Sándor Á, Shum SB, Ferguson R, De Liddo A, Whitelock D (2015) Correlations between automated rhetorical analysis and tutors' grades on student essays. In: Proceedings of the fifth international conference on learning analytics and knowledge. ACM, pp 355–359

Sinclair J, Kalvala S (2015) Engagement measures in massive open online courses. In: International workshop on learning technology for education in cloud. Springer International Publishing, pp 3–15

Slade S, Prinsloo P (2013) Learning analytics ethical issues and dilemmas. Am Behav Sci 57 (10):1510–1529

Softic S, Taraghi B, Ebner M, De Vocht L, Mannens E, Van de Walle R (2013) Monitoring learning activities in PLE using semantic modelling of learner behaviour. Human factors in computing and informatics. Springer, Berlin, pp 74–90

Strang KD (2016) Beyond engagement analytics: which online mixed-data factors predict student learning outcomes? Education and information technologies, pp 1–21

Swenson J (2014) Establishing an ethical literacy for learning analytics. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 246–250

Tervakari AM, Marttila J, Kailanto M, Huhtamäki J, Koro J, Silius K (2013) Developing learning analytics for TUT Circle. Open and social technologies for networked learning. Springer, Berlin, pp 101–110

Tseng SF, Tsao YW, Yu LC, Chan CL, Lai KR (2016) Who will pass? Analyzing learner behaviors in MOOCs. Res Pract Technol Enhanced Learning 11(1):1

Vahdat M, Oneto L, Anguita D, Funk M, Rauterberg M (2015) A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. In: Design for teaching and learning in a networked world. Springer International Publishing, pp 352–366

van Barneveld A, Arnold KE, Campbell JP (2012) Analytics in higher education: establishing a common language. EDUCAUSE Learning Initiative 1:1–11

Vozniuk A, Holzer A, Gillet D (2014) Peer assessment based on ratings in a social media course. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 133–137

Westera W, Nadolski R, Hummel H (2013) Learning analytics in serious gaming: uncovering the hidden treasury of game log files. In: international conference on games and learning alliance. Springer International Publishing, pp 41–52

Wise AF (2014) Designing pedagogical interventions to support student use of learning analytics. In: Proceedings of the fourth international conference on learning analytics and knowledge. ACM, pp 203–211

Wolff A, Zdrahal Z, Nikolov A, Pantucek M (2013) Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In: Proceedings of the third international conference on learning analytics and knowledge. ACM, pp 145–149

Wu IC, Chen WS (2013) Evaluating the practices in the e-learning platform from the perspective of knowledge management. Open and social technologies for networked learning. Springer, Berlin, pp 81–90

Yasmin D (2013) Application of the classification tree model in predicting learner dropout behaviour in open and distance learning. Dis Educ 34(2):218–231