

A Review of Learning Analytics in Educational Research

Kam Cheong Li^(✉), Hoi Kuan Lam, and Sylvia Siu-Yin Lam

The Open University of Hong Kong, Hong Kong SAR, China
{kcli, hlam, ssylam}@ouhk.edu.hk

Abstract. An increasing amount of research work is now focused on the topic of learning analytics. Gathering the facts from the literature is important for identifying issues and features in ‘the state of the art’ and the way forward. This paper aims to examine the research literature published in the last five years and construct a systematic review of learning analytics. A document analysis method is used to classify the research work in 51 articles, selected from the Web of Science, which report educational research studies using a learning analytics approach. The details of the studies are categorised according to their research questions or objectives, the methodology (such as input data, techniques used or software tools for data collection and analysis) and the findings. This paper offers an overview of the emerging field of learning analytics, provides a foundation for exploring this promising area of educational research, and identifies a series of future challenges.

Keywords: Learning analytics · Education research · Systematic review

1 Introduction

Learning analytics (LA) is an emerging field in the education sector. It focuses specifically on the learning process [1] and involves the use of big data techniques to capture, model and predict the behaviours of diverse target groups from a massive volume of unstructured data. At academic institutions, LA is used to examine the relevant data on students and instructors at a micro-level which target individual learners and the courses taken in order to understand student performance and promote student success [2]. With the sophisticated analytic tools and techniques of LA, student performance and learning outcomes can be improved by enhanced targeting of support and intervention, thus promoting learning and education [3]. The study and advancement of LA involves the development, usage and integration of new processes and tools in order to improve the practice of learning and teaching for individual students and instructors.

2 Background

In communities of educators, LA and educational data mining (EDM) form two research areas oriented towards the inclusion and exploration of big data capabilities in education for gaining insights into the learning activities of learners [4]. EDM is an area for “developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyse

due to the enormous volume of data within which they exist” [5]. LA is defined as “the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs” [6]. There are other definitions of the term LA which are different in some details, but the definitions share an emphasis on converting educational data into useful actions to foster learning [7].

Although EDM and LA share the common goal of gaining insights into learners’ activities, they are different in their origins, techniques, fields of emphasis and types of discovery [5, 7, 8]. Nonetheless, the two research areas are complementary [4]. The research results on EDM do not focus on empirical evidence but on the objectives, methods, processes and tools for knowledge discovering. LA, on the other hand, adopts a holistic approach when seeking insights into the learning processes. An overview of LA argues that teachers should engage with LA for richer conceptions of learning and improvements in teaching [9]. Another study supplements the insights for students, stating that “receiving information about their performance in relation to their peers or about their progress in relation to their personal goals can be motivating and encouraging” [1].

The concepts and methods of LA are drawn from a variety of related fields [7]. It is “an area of research related to business intelligence, web analytics, academic analytics, action analytics and predictive analytics” [4]. LA is also a field in which several related areas of research in technology-enhanced learning converge, including academic analytics, action research, EDM, recommender systems and personalized adaptive learning [7]. The more concrete examples of LA practice comprise predictive modelling, social network analysis (SNA), usage tracking, content analysis and semantic analysis, and recommendation engines [9].

Given the common emphasis on converting educational data to support the learning process and foster learning, as well as the different practices of LA, there is however no concrete theoretical model or framework of LA in the literature. As Clow states, “Learning analytics is not so much a solid academic discipline with established methodological approaches as it is a ‘jackdaw’ field of enquiry, picking up ‘shiny’ techniques, tools and methodologies This eclectic approach is both a strength and a weakness: it facilitates rapid development and the ability to build on established practice and findings, but it — to date — lacks a coherent, articulated epistemology of its own” (pp. 685–686) [9]. In Papamitsiou and Economides’s literature review of empirical evidence of LA and EDM, they assert that “The motivation for this review derived from the fact that empirical evidence is required for theoretical frameworks to gain acceptance in the scientific community.... Consequently, there was a need to supply the audience with an accredited overview” (p. 50) [4]. Based on the LA reference model, Chatti and others reviewed relevant studies in LA applications and mapped the studies onto the four dimensions of the model, namely data and environments, stakeholders, objectives and methods [7]. The review, however, was confined to two years, 2010 and 2011. A search in the relevant literature did not find any review of empirical evidence of LA for a longer period, which motivated us to produce a critical review of empirical studies of LA over a five-year period to indicate the extent of maturity and deployment of LA applications for useful actions to foster learning.

3 Methodology

3.1 Aim and Objectives

Though LA has demonstrated its potential as a promising research area in educational technology, only limited systematic literature reviews have been carried out on the topic. This paper, which aims to collect and summarize information derived from the literature about the applications of LA in educational research, addresses the following research questions:

- 1 What kinds of educational research have been conducted using an LA approach?
- 2 What kinds of data have been used for educational research using an LA approach?
- 3 What kinds of techniques/software tools are available for educational research using an LA approach?
- 4 What kinds of key findings are observed from educational research using an LA approach?

3.2 Data Collection and Analysis Methods

This paper aims to examine the research literature about LAs published in scholarly journals. Studies from this research have been chosen by accessing electronic sources. The literature is limited to studies published in the last five years, between 2011 and 2015, in the Web of Science from international databases. The data sources for this study are summarized in Table 1(A). The Web of Science includes such journals as the *Journal of the Learning Sciences*, *Computers & Education*, *The Internet and Higher Education*, the *International Journal of Computer-Supported Collaborative Learning*, *Learning Media and Technology*, the *British Journal of Educational Technology*, *Sport Education and Society*, *IEEE Transactions on Learning Technologies*, the *Australasian Journal of Educational Technology*, the *Journal of Geography in Higher Education*, *Educational Technology and Society*, *Distance Education*, *Teaching in Higher Education*, the *International Review of Research in Open and Distance Learning*, and *Culture, Education and Communication*. While searching, the word ‘LA’ has been used as the search topic and the search category has been limited to ‘education educational research’. The resulting search gained access to a total of 51 studies, in which 47 were articles, two were reviews and two were editorials. In this study, document analysis has been used to examine each article and the content was identified through objective, systematic and quantitative categorization [10]. Through document analysis, the information extracted from the selected literature has been examined and revised using a particular encoding system and has been used as collected data [11]. Subsequently, the data gathered by document analysis have been made into content analysis, according to the mathematical representation of the data based on the characteristics observed [12]. The literature has been examined carefully and categorized into three main criteria. These criteria within the framework of the research approach were (1) the research question or objective of the studies; (2) the methodology used in the studies; and (3) the key findings of the studies. The second criterion — the methodology used — was further divided into five sub-criteria. They were (2.1) the data source, i.e. the kind of system in which the data were gathered, managed and used for the analysis; (2.2) stakeholders,

i.e. the participant(s) targeted by the analysis; (2.3) study group, i.e. the characteristics of the participants; (2.4) instrument(s), i.e. the technique(s) used to perform the analysis of the collected data; and (2.5) course or field of the study, i.e. the area that the study applied to. Comparable criteria have been used before for the same purposes [7, 13], but there have also been some different criteria included in this paper. The three main criteria and the five sub-criteria have been considered within the scope of this research. The criteria for examination in this study are summarized in Table 1(B).

Table 1. Data source and criteria for examination in this study

(A) Data source	
Database:	Web of Science
Search topic:	Learning analytics
Search category:	Education educational research
Search time-span:	2011–2015 (five years)
Search result:	51 pieces of literature in total (including two reviews, 47 articles, and two editorials)
(B) Criteria for the examination of the articles used	
(1) The research question or objective of the studies	
— Monitoring and analysis; prediction and intervention; assessment and feedback; adaptation; personalization and recommendation; and reflection	
(2) The methodology used in the studies	
(2.1) Data source: Where did the educational data come from?	
— Closed/Protected, e.g. a learning management system vs open/distributed, e.g. personal learning environment	
(2.2) Stakeholder: Who is/are the participant(s)?	
— Student; teacher; educational institution; researcher; and system designer	
(2.3) Study group: What is/are the characteristic(s) of the participant(s)?	
— Primary school; secondary school; higher education	
(2.4) Instrument: What is/are the technique(s) used to perform the analysis of the collected data?	
— Survey/questionnaires; statistics; non-statistics; information visualization; data mining; social network analysis; content analysis; natural language processing; machine learning; group concept mapping; pattern information analysis; and ethnographic analysis	
(2.5) Course or field of study: To which area does the study apply?	
— Education technology; science, technology, engineering and mathematics; geographical education; health and physical education; educational research; computer science; education publications; humanities; media literacy education; medical education and digital image processing	
(3) The key findings of the studies	
— Learning outcomes (measured as positive, negative and neutral)	

4 Results

LA has a huge potential for supporting learning, teaching and education, and the number of publications on LA research has grown rapidly in the last few years. In this paper, the selected literature on LA for the research has been analysed for three main criteria and five sub-criteria as described above in Table 1(B), and the results can be summarized as follows.

4.1 Distribution of the Research Question or Objectives

There are many objectives in the selected literature and they have been examined one by one. Similar topics have been combined, which include monitoring and analysis, prediction and intervention, assessment and feedback, adaptation, personalization and recommendation, and reflection. The distribution of the studies on research questions or objectives is summarized in Table 2.

Table 2. Distribution of literature by research questions or objectives

Research question/Objective	Frequency (%)
Monitoring and analysis	10 (19.6 %)
Prediction and intervention	6 (11.8 %)
Assessment and feedback	8 (15.7 %)
Adaptation	9 (17.6 %)
Personalization and recommendation	3 (5.9 %)
Reflection	15 (29.4 %)

Table 2 shows that the most frequent objectives in the selected literature are on reflection (29.4 %), and monitoring and analysis (19.6 %). Fewer objectives aim for adaptation (17.6 %), assessment and feedback (15.7 %), prediction and intervention (11.8 %), and personalization and recommendation (5.9 %). This answers our research question 1 — ‘What kinds of educational research have been conducted using an LA approach?’ — by research question or objective.

4.2 Distribution of Data Source

The LA tools that have been proposed in the literature selected use different data sources. We classified the data sources into closed/protected [e.g. learning management system (LMS)] and open/distributed [e.g. personal learning environment (PLE)]. The distribution of the studies on where the educational data came from is summarized in Table 3.

Table 3. Reviewed literature by distribution of data source

Data source	Frequency (%)
Closed/Protected, e.g. LMS	23 (45.1 %)
Open/Distributed, e.g. PLE	28 (54.9 %)

Table 3 illustrates that studies from the chosen literature are 54.9 % from open or distributed sources and 45.1 % from closed or protected sources. The open or distributed sources include literature; Elgg®, the social networking engine; computer-supported collaborative learning (CSCL); web-based systems (such as wikis, learning and content management systems, forums, academic portals, repositories); and massive open online courses (MOOC). The closed or protected data sources include Equella; computer-assisted curriculum analysis; design and evaluation (CASCADE); virtual field trip (VFT); and QuesTInSitu — the Game, LOCO-Analyst and Blackboard. This answers our research question 2 — ‘What kinds of data have been used for educational research using an LA approach?’ — by the distribution of data source.

4.3 Distribution of Stakeholders

The stakeholders who participated in studies in the selected literature include students, teachers, educational institutions, researchers and system designers. The distribution of the studies on the participants is summarized in Table 4.

Table 4. Reviewed literature by distribution of stakeholders

Participant	Frequency (%)
Students	17 (23.6 %)
Teachers	21 (29.2 %)
Educational institutions	19 (26.4 %)
Researchers	12 (16.7 %)
Systems designers	3 (4.2 %)

Table 4 shows that most of the studies have targeted teachers (29.2 %), educational institutions (26.4 %) and students (23.6 %). Fewer studies have involved researchers (16.7 %) and system designers (4.2 %). This also answers our research question 2 — ‘What kinds of data are used for educational research using an LA approach?’ — according to the participants in the studies.

4.4 Distribution of Study Group

The studies selected have been classified by study group according to the participants’ characteristics, which include primary school, secondary school, and higher education.

The distribution of the studies on study group according to participants' characteristics is summarized in Table 5.

Table 5. Reviewed literature by distribution of participants' characteristics

Study group	Frequency (%)
Primary school	0 (0 %)
Secondary school	6 (17.1 %)
Higher education	29 (82.9 %)

Table 5 illustrates that the study groups in the studies are concentrated heavily in higher education (82.9 %). There are fewer studies at the level of secondary schools (17.1 %) and none at the primary school level (0 %). This also answers our research question 2 — 'What kinds of data have been used for educational research using an LA approach?' — by study group according to participants' characteristics in the studies.

4.5 Distribution of Instruments

The selected studies have been classified according to the instruments used, which include surveys/questionnaires, statistics, non-statistics, information visualization (IV), data-mining (DM), SNA, content analysis, natural language processing (NLP), machine learning, group concept mapping, pattern information analysis and ethnographic analysis. Note that some studies applied a variety of methods and can therefore be found in multiple categories. The distribution of the studies on instruments used is summarized in Table 6.

Table 6. Reviewed literature by distribution of instruments

Technique or software tool	Frequency (%)
Surveys/Questionnaires	9 (14.3 %)
Statistics	3 (4.8 %)
Non-statistics	18 (28.6 %)
Data mining	8 (12.7 %)
Machine learning	3 (4.8 %)
Information visualization	9 (14.3 %)
Social network analysis	6 (9.5 %)
Content analysis	3 (4.8 %)
Natural language processing	1 (1.6 %)
Group concept mapping	1 (1.6 %)
Pattern information analysis	1 (1.6 %)
Ethnographic analysis	1 (1.6 %)

As can be seen in Table 6, the most used LA techniques in the literature reviewed take advantages of information retrieval technologies with classical tools, such as non-statistics (28.6 %), surveys or questionnaires (14.3 %), IV (14.3 %), DM (12.7 %), machine learning (12.7 %) and statistics (4.8 %). Other techniques, such as SNA (9.5 %), content analysis (4.8 %), NLP (1.6 %), group concept mapping (1.6 %), pattern information analysis (1.6 %) and ethnographic analysis (1.6 %) are also employed in the studies. This answers our research question 3 — ‘What kinds of techniques/software tools are available for educational research using an LA approach?’ — on the instruments used.

4.6 Distribution of Course of Research or Field of Study

The studies chosen have been classified according to application courses and their fields, which include education technology, science, technology, engineering and mathematics (STEM), geographical education, health and physical education, educational research, computer science, education publications, humanities, media literacy education, medical education and digital image processing. The distribution of the studies on the course of research or field of study is summarized in Table 7.

Table 7. Reviewed literature by distribution of course or field of study

Course/Field of study	Frequency (%)
Education technology	21 (38.2 %)
Science, technology, engineering and mathematics (STEM)	10 (18.2 %)
Geographical education	1 (1.8 %)
Health and physical education	1 (1.8 %)
Educational research	11 (20.0 %)
Computer science	5 (9.1 %)
Education publications	1 (1.8 %)
Humanities	1 (1.8 %)
Media literacy education	1 (1.8 %)
Medical education	2 (3.6 %)
Digital image processing	1 (1.8 %)

Table 7 shows that studies in the selected literature focus on education technology (38.2 %) and educational research (20.0 %) on LA. The effectiveness studies oriented to different courses are mainly on STEM (18.2 %), computer science (9.1 %) and medical education (3.6 %). Fewer studies are concerned with courses such as geographical education (1.8 %), health and physical education (1.8 %), education publications (1.8 %), humanities (1.8 %), media literacy education (1.8 %) and digital image processing (1.8 %). Again,

this answers our research question 2 — ‘What kinds of data have been used for educational research using an LA approach?’ — by the courses or fields of the studies.

4.7 Distribution of the Key Findings

The studies chosen have been classified according to positive, negative and neutral learning outcomes, the distribution of which is summarized in Table 8.

Table 8. Reviewed literature by distribution of learning outcomes

Learning outcome	Frequency (%)
Positive	45 (88.2 %)
Negative	2 (3.9 %)
Neutral	4 (7.8 %)

In Table 8, it can be seen that the majority of the studies have a positive learning outcome (88.2 %) within the scope of the research. Meanwhile, the learning outcomes in two studies are negative (3.9 %), and in another four they are neutral (7.8 %). This answers our research question 4 — ‘What kinds of key findings are observed for educational research using an LA approach?’ — by learning outcomes.

5 Discussion and Conclusion

The literature analysed in this study has been chosen from the Web of Science which is an online subscription-based scientific citation indexing service. It provides a comprehensive citation search by accessing multiple databases that reference cross-disciplinary research and allows an in-depth exploration of specialized sub-fields within an academic or scientific discipline. Therefore, the selected literature on the topic LA in the field of ‘educational education research’ in this study is highly relevant and comes from journals with an impact factor ranging from 0.35 to 3.26. This is already evidence that LA, which involves large amounts of data in combination with information retrieval technologies, has substantial potential for use in education [14]. The novel information retrieved from LA can support individual learning as well as organizational knowledge management [15]. Research on the application of LA in education has been increasing since 2011 and this has been sustained up to the present.

The advantages of LA are that it reveals and translates the educational data from unknown to meaningful information and prepares it for students, teachers and educational institutions [13]. The objective which is applied most in the literature reviewed in this study is reflection (29.4 %) which has been distinguished as a fundamental objective in LA since 2009 [16]. Reflection is a process involving quantifying oneself from one’s own performance for better learning outcomes. The second most common objective applied in the selected literature is monitoring and analysis (19.6 %) which, by comparing information on and interactions with students, can offer new perceptions of

both learners and organizations in terms of effectiveness and efficiency. The third most frequent applied objective in the literature reviewed is adaptation (17.6 %), which adaptively articulates learners to the next move by consolidating learning resources and instructional activities according to individual learner's needs [7].

As there is a shift in focus from centralized learning systems to open learning environments, the use of closed/protected data sources as a dominant trend has changed. Our findings show that more open/distributed data sources (54.9 %) have been used in the studies in the selected literature, as compared to closed/protected data sources (45.1 %). The closed/protected data sources, such as LMS, have been dominant since the emergence of LA, while open/distributed data sources, such as PLE, have grown considerably in recent years; and the use of closed/protected and open/distributed data sources have become fairly balanced.

LA studies of pedagogical issues undoubtedly involve students and teachers as stakeholders. Traditionally, the investigation of students' behaviours and activities has been one of the main focuses in LA research. These studies emphasize the generation of student-centred feedback by tracking users' data from learning systems, but much less research was concerned with educator-centred feedback. However, recently, there has been a tendency for much more stress to be put on stakeholders other than students. Our findings have shown that the majority of the studies in the literature chosen have targeted teachers (29.2 %), educational institutions (26.4 %) and students (23.6 %), suggesting that educator-centred studies have been increasing. The involvement of stakeholders, such as researchers and system designers, has provided a more comprehensive view of information using LA.

The study group in the LA research in the literature reviewed has been focused on the level of higher education (82.9 %), with fewer studies at the secondary school level (17.1 %) and none on the primary school level. This phenomenon may be due to the fact that the research subjects are generally those within the age range for higher education who can take advantages of the technology by having adequate learning skills. Stakeholders, such as students at university level, fit these criteria well and are perfect subjects for researchers who are most likely to also be working in higher education institutions.

Different techniques or software tools can be applied in the development of education applications that support the objectives of educational stakeholders. LA takes advantage of information retrieval technologies that can contribute tailored information support systems to the stakeholders on demand, and can be applied to a vast variety of field of study [14]. It is clear from our findings that the techniques or software tools used range from classical LA tools to the latest advanced technological tools, such as the study conducted through mobile applications by Melero et al. [17]. In this sense, there would be no boundary for the application courses or fields of those studies, as shown in our findings on a wide variety of courses, ranging from humanities and medical to STEM. Nonetheless, education technology courses were the major field as technological advances play a critical role in the development of LA research.

The key findings on LA examined in terms of learning outcomes within the scope of research were mostly constructive. More than 80 % of the studies indicated positive learning outcomes, suggesting that LA as a field has strengthened learning, teaching and pedagogical decision-making. However, two studies showed that, regardless of how

powerful and promising LA is as a technological advance in guiding and appraising the educational progress, technologies alone are not enough for seeing the whole picture [18, 19]. It is sensible to take into consideration human beings who are properly trained, determined, and dedicated to education — such as teachers, system designers, policy administrators and maybe parents — in order to complete the picture as a whole.

Research in the field of LA has been booming in the last five year, but LA is still at the infant stage in its development. Universities should take careful note of its advances and potential for use, together with conventional methods of student support, to achieve substantial improvements in the practice of higher education. This active research area will continue to contribute valuable pieces of work to the development of powerful and mostly accurate learning services for both learners and teachers.

References

1. Siemens, G., Long, P.: Penetrating the fog: analytics in learning and education. *EDUCAUSE Rev.* **46**(5), 30–40 (2011)
2. Learning Analytics: Definitions, Processes and Potential. <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
3. van Barneveld, A., Arnold, K.E., Campbell, J.P.: Analytics in higher education: establishing a common language. *EDUCAUSE Learn. Initiative.* **1**, 1–11 (2012)
4. Papamitsiou, Z., Economides, A.: Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Educ. Tech. Soc.* **17**(4), 49–64 (2014)
5. Romero, C., Ventura, A.: Data mining in education. *WIREs Data Min. Knowl. Disc.* **3**, 12–27 (2013)
6. Proceedings of the 1st International Conference on Learning Analytics and Knowledge. <https://tekri.athabascau.ca/analytics/>
7. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H.: A reference model for learning analytics. *Int. J. Technol. Enhanced Learn.* **4**(5–6), 318–333 (2012)
8. Siemens, G., Baker, R.: Learning analytics and educational data mining: towards communication and collaboration. In: Buckingham Shum, S., Gasevic, D., Ferguson, R. (eds.) *2nd International Conference on Learning Analytics and Knowledge*, pp. 252–254. ACM, New York (2012)
9. Clow, D.: An overview of learning analytics. *Teach. High. Educ.* **18**(6), 683–695 (2013)
10. Bauer, M.W.: Classical content analysis: a review. In: Bauer, M.W., Gaskell, G. (eds.) *Qualitative Researching with Text, Image and Sound*, pp. 131–151. Sage Publication, London (2003)
11. Çepni, S.: Araştırma ve proje çalışmalarına giriş (3. Baskı), Methods and classifications used in educational research. Celepler Publisher, Trabzon (2007)
12. Karasar, N.: Bilimsel araştırma yöntemi [Research Methods], 11. Nobel Publisher, Ankara (2002)
13. Greller, W., Drachsler, H.: Translating learning into numbers: a generic framework for learning analytics. *Educ. Tech. Soc.* **15**(3), 42–57 (2012)
14. Johnson, L., Smith, R., Willis, H., Levine, A., Haywood, K.: The 2011 Horizon Report. The New Media Consortium, Austin (2011)
15. Butler, D.L., Winne, P.H.: Feedback and self-regulated learning: a theoretical synthesis. *Rev. Educ. Res.* **65**, 245–281 (1995)

16. Know Thyself: Tracking Every Facet of Life, from Sleep to Mood to Pain. *Wired Magazine*. http://www.wired.com/medtech/health/magazine/17-07/lbnp_knowthyself?currentPage=all
17. Melero, J., Hernández-Leo, D., Sun, J., Santos, P., Blat, J.: How Was the Activity? A Visualization Support for a Case of Location-Based Learning Design. *Br. J. Educ. Technol.* **46**(2), 317–329 (2015)
18. Macfadyen, L.P., Dawson, S.: Numbers Are Not Enough. Why E-learning Analytics Failed to Inform an Institutional Strategic Plan. *Educ. Technol. Soc.* **15**(3), 149–163 (2012)
19. Spector, J.M.: Emerging educational technologies and research directions. *Educ. Technol. Soc.* **16**(2), 21–30 (2013)