Wu, Y.-T., Chang, M., Li, B., Chan, T.-W., Kong, S. C., Lin, H.-C.-K., Chu, H.-C., Jan, M., Lee, M.-H., Dong, Y., Tse, K. H., Wong, T. L., & Li, P. (Eds.). (2016). *Conference Proceedings of the 20th Global Chinese Conference on Computers in Education 2016*. Hong Kong: The Hong Kong Institute of Education.

Topic Structure and Visualization Analysis of Learning Analytics Researches

Xinger LU, Junce FANG, Zhinan HUANG, Menglu JIANG, Yanyan LI*
R&D Center for Knowledge Engineering, School of Education Technology, Beijing Normal University, Beijing
*fccg0121@126.com

Abstract: *As an important part of educational data mining, learning analytics has drawn extensive attention among international scholars since it was proposed. To further explore the research topic of learning analytics, this paper, which bases on the data containing 674 documents on learning analytics between 2010 and 2015 from Web of Science, explores and reveals the international research status about learning analytics. In the process of that, we adopt a variety of analysis tools and visualization technology to analyze the high cited documents, the distribution of core authors, high prolific institutions and high frequency keywords. Besides, on the basis of high frequency keywords dissimilarity matrix, we further clarify the research scope by the means of cluster analysis. Furthermore, by mapping out the strategic diagram graph, the trend of the development orientation of learning analytics is further clarified in order to provide reference and suggestions for international research and practice on learning analytics.*
Keywords: learning analytics, co-word analysis, visualization, cluster analysis

1. Introduction

With the rapid development of information technology, human society has entered the era of big data. As an important part of educational data mining, learning analytics helps to understand and optimize learning process and environment, through measurement, collection, analysis and reporting data of learners and their learning situation (Ferguson, 2012). Learning analytics has drawn extensive attention among international scholars, and they have studied learning analytics from different angles, which reached different conclusions. Among them, Siemens (2009) put forward TEKL Model, which changed traditional classroom teaching, and further proposed the specific learning and analyzing process model on the basis of it. Wolfgang and Hendrick (2012) summarized discussed topics in the forum of research and analysis of study by using Ritchie's general morphological analysis, then formed the learning and analysis elements model. The scholars from Korean Dankook University combining on-line interaction visualization technology with their research, put forward the multidimensional analysis method which analyzes the learning data by various dimensions (Kim & Lee, 2012). Generally speaking, the international research of learning analytics is gradually mature. In order to clearly grasp the research status of learning analytics, this paper researches on the existing literature by the means of scientific metrology method.

The data for this paper contains 674 documents on learning analytics published between 2010 and 2015 which are downloaded from Web of Science. By using Bicomb word frequency analysis software, CiteSpace citation analysis tools, SCI2 knowledge map building tools, SPSS statistical analysis software, etc., drawing co-citation network map, social network analysis map based on the co-word of keywords, dendrogram, strategic diagrams and so on. We present the history, current situation, research area and development trend of learning analytics from different perspectives, to provide reference for international study of learning analytics.

2. Methods

Knowledge graph combines theory science, such as mathematics, graphics, information science, etc., with metrology methods, which uses visual images to show knowledge resources, and also analyzes, builds and shows the relationship between various knowledge. The main methods used in knowledge map are co-word analysis, citation

analysis, social network analysis, etc.

3. Findings

3.1. Citation Analysis

3.1.1. Analysis of High Cited Paper

High cited literatures are those whose citations rank at the top of disciplines in a statistical period. We import downloaded literatures into CiteSpace; set up timespan period to 2010-2015; set up to display the highest frequency of reference at top 50 within each slice as a result, and visualize them. Figure 1 below shows the high cited literature visualization diagrams, the cited frequencies are greater than or equal to 6.

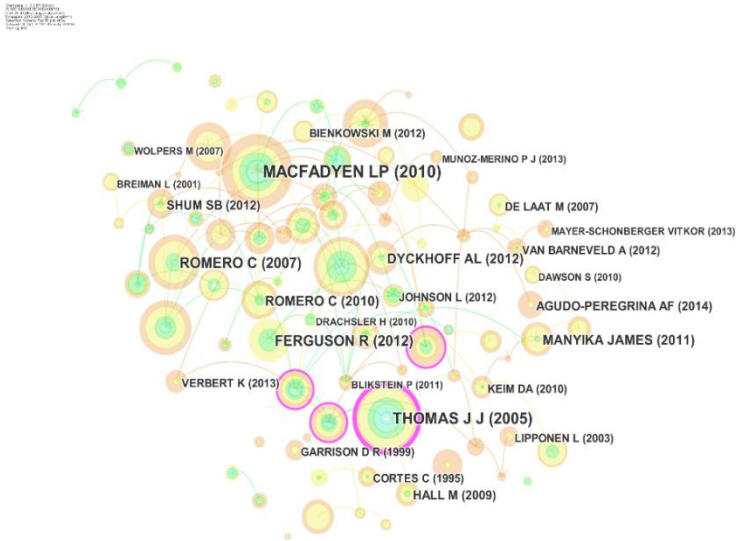

Figure 3. High cited literature (cited frequency>=6)

As can be seen from Figure 1, the paper James J. Thomas published in 2005 called "Illuminating the Path: The Research and Development Agenda for Visual Analytics" has a higher cited frequency. This thesis pointed out a visual analysis includes analytical reasoning techniques, visualization techniques and interactive performance technology, data presentation and conversion technology, and products supporting, demonstration and dissemination technology, which laid a theoretical guidance for learning analytics visualization development. "Mining LMS data to develop an 'early warning system' for educators: A proof of concept" written by Leah P.Macfadyen in 2010 put forward that, by collecting and analyzing data from both study platforms and non-study platforms, the early warning system enabled the learning management system to show students' participation as well as their achievements acquired in learning courses. Kalina Yacef published a paper called "The state of educational data mining in 2009: A review and future visions". By sorting out the existing relevant researches on educational data mining, she drew a conclusion that the amount of published papers and public educational database on Educational Data Mining and educational data conferences has increased, while the barriers to become an educational data mining researcher on on-line course built-in tools has decreased. Analyzing existing high cited literature of learning analytics helps to grasp the development direction of this field.

3.1.2. Analysis of High Cited Institution

We use CiteSpace for statistical analysis, the result is showed as figure 2 that mainly high cited institutions, including Universidad Carlos III de Madrid, University of Sydney, Athabasca University, IBM Corp, Open University of Netherlands, etc. As can be seen, the total number of nodes is 205, the number of sides is 45, and the density is only 0.0022, which shows the research of learning analytics is still under development and immature. In addition, some small cooperative groups among research institutions has been formed, such as Athabasca University, Dalhousie University

and Simon Fraser University, but wider range of cooperation has not yet been constructed.

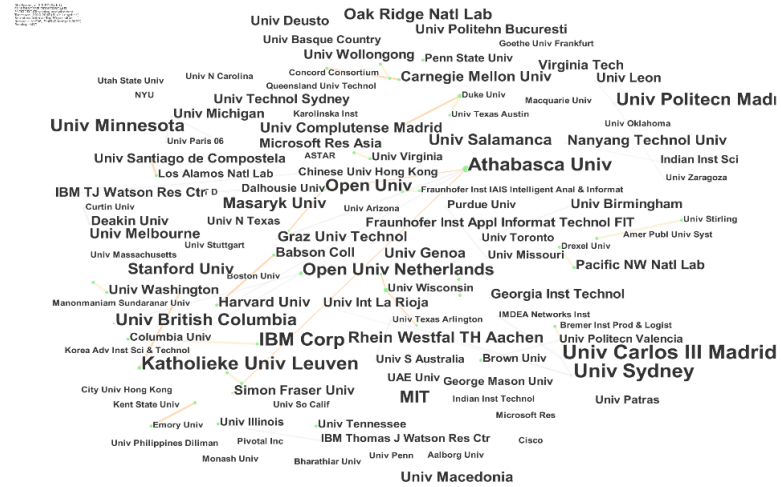

Figure 4. High cited institution

3.1.3. Analysis of High Cited Journal

By analyzing the amount of citations, we can further insight into the quality as well as other information of journals and papers. We import downloaded data into CiteSpace, and set node type as Cited Journal, the visual analysis result of high cited journals is showed as figure 3. As we can see, higher cited journals and books of learning analytics are Lecture Notes in Computer Science, Expert System Application, British Journal of Educational Psychology, Computers & Education, Educational Technology and Society, etc.

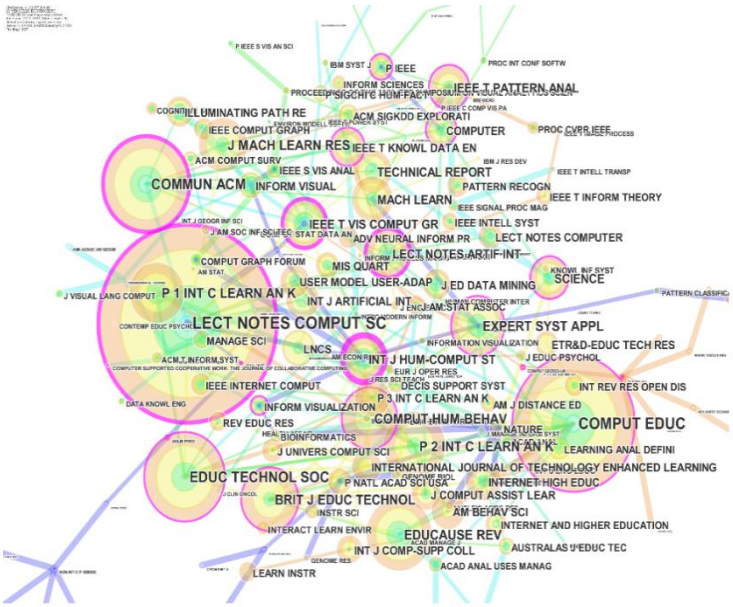

Figure 5. High cited journal

3.2. Co-word Analysis

3.2.1. High Frequency Keyword Analysis

We import the 674 documents into Bicomb and analyze high frequency keywords. Considering some of them express the same meanings in different forms, some have the same forms but are different in uppercase and lowercase letters, we plan to conduct data merging and other semantic processes. For example, “Learning Analytics” and “learning analytics” are unified into “learning analytics”. We reached 2021 high frequency keywords as a result, 556 of which appear more than 2 times with 56.7% cumulative frequency. And 36 of them appear more than 5 times with 1.78%

cumulative frequency. A part of the list of high frequency keywords is shown in Table 1.

Table 1. High frequency keywords

No.	Keyword	Freq.	No.	Keyword	Freq.	No.	Keyword	Freq.
1	learning analytics	257	8	Prediction	26	15	personalized learning	9
2	Big Data	81	9	Clustering	22	16	Mobile learning	8
3	Machine Learning	59	10	Social Networks	17	17	collaboration	8
4	Educational data mining	57	11	Assessment	17	18	Models	8
5	Visualization	52	12	MOOCS	13	19	online learning	7
6	Visual analytics	40	13	Social media	13	20	Moodle	7
7	E-learning	27	14	CSCL	11	21	Hadoop	7

3.2.2. Co-occurrence Matrix of High Frequency Keyword

Co-occurrence matrix of high frequency keywords is produced by Bicomb. In order to clearly show the intimacy between keywords, this paper imports 36 high-frequency keywords into SCI2 to generate the co-occurrence network diagram, as shown in figure 4. Results show that, there are 36 nodes in total, with 110 edges among them. The minimum weights of figure node are 1, while the maximum are 21, the network average total degree is 6.11 and the density is 0.0873. Among these words, Big data, Educational data mining, MOOCs, Moodle, Visual analytics, etc. have a high degree of centrality. The result indicates that big data mining and visual analysis based on MOOCs and other online education platform are important parts of learning analytics research.

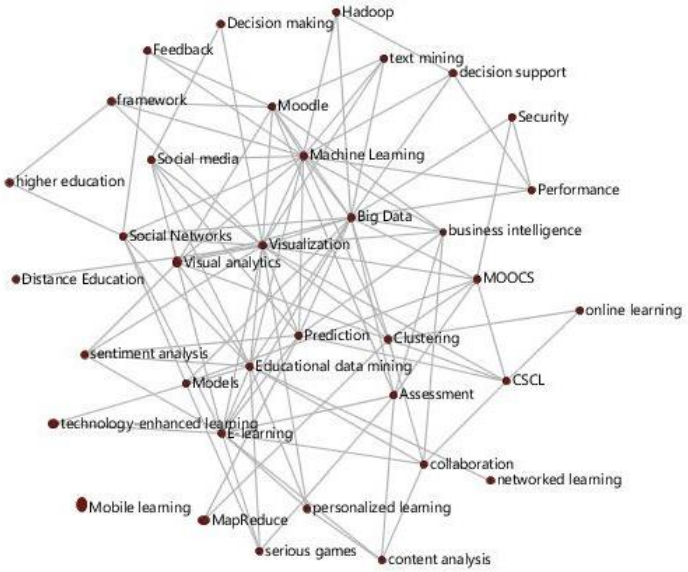

Figure 6. Co-occurrence network diagram of high-frequency keywords

3.3. Topic Clustering

To further explore the research topic of learning analytics, we import the dissimilarity matrix into UCINET6 for keyword clustering, and form a class of keywords having close relationship to represent a particular field. Figure 5 shows the topic clustering graph.

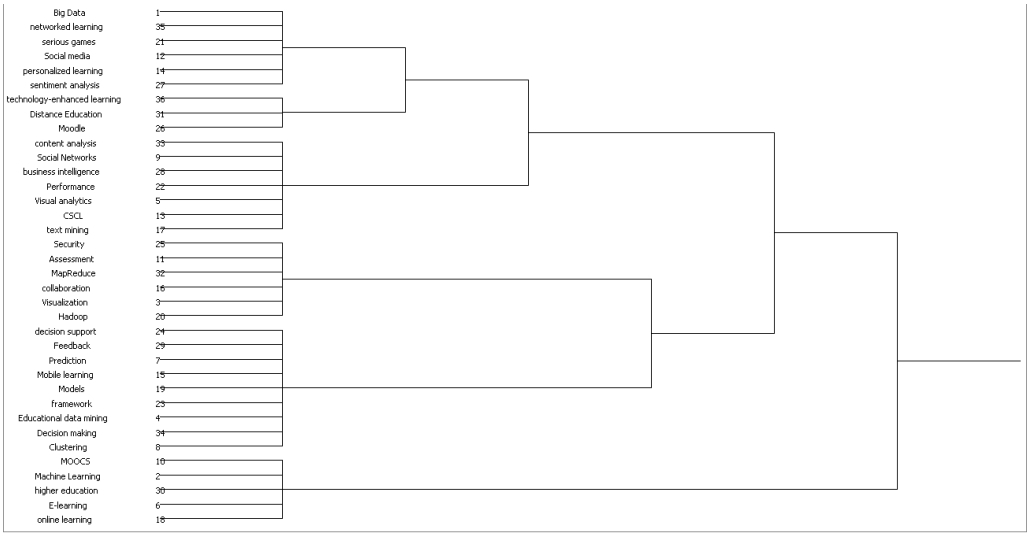

Figure 7. Topic clustering

According to the results of cluster analysis, we can divide the keywords of learning analytics into five class groups. As shown in Table 2.

Table 2. Subject classification

No.	Class group	Keywords
1	Framework model research in learning analytics	framework, Prediction, Models, Clustering, decision support, Decision making, Feedback, Mobile learning, Educational data mining
2	Learning environment research	Big Data, networked learning, serious games, Social media, personalized learning, sentiment analysis, technology-enhanced learning, Distance Education, Moodle
3	Tools and method research in learning analytics	content analysis, Visual analytics, text mining, Social Network, business intelligence, Performance, CSCL
4	Technical support research in learning analytics	MapReduce, Hadoop, collaboration, Visualization, Assessment, Security
5	Application research in learning analytics	MOOCS, Machine learning, higher education, E-learning, online learning

3.3.1. Framework Model Research in Learning Analytics

Framework model research is the foundation of learning analytics research, which provides the overall framework and main factors of learning analytics. Although there are a variety of constructions of Framework model theory in the field of learning analytics, the relationship analysis among students, learning and curriculum are inevitable cores of design. Under the guidance of the above philosophy, multi-information analysis framework constructs an analysis system with feedback process which combines data, analytical tools and environment. Such analytical links are achieved through predicting learning outcomes and using cluster analysis technology to integrate the analytical data of online learning and mobile learning (Ifenthaler & Widanapathirana, 2014). Different from educational data mining, learning analytics focuses more on social intervention, it aims to optimize the learning process and learning environment by influencing decision-making results and further providing decision support. (Siemens & d Baker, 2012). This kind of frameworks are diverse, which, to a certain extent, guide the theoretical research. But there is still a long way to prove them by practice.

3.3.2. Learning Environment Research

Learning environment research is one of the main objects of study of learning analytics. Through the analysis of the learning environment, we can provide solutions for future decision on environment enhancement. International scholars held the view that, to a large extent, learning analytics means applying "big data" analytics to the educational field (Ifenthaler & Widanapathirana, 2014). In current pervasive computing era, it's been a common phenomenon to apply electronic equipment to learning situated context, and this technology-enhanced learning promotes Learning Analytics to further develop. With the influences of ubiquitous networks, such as social media and the Internet on study, a lot of learning information has been reserved through distance education, Moodle and other forms of network learning and online learning, which become main sources of the data. And through emotional analysis of these articles, students' learning states and attitudes are obtained and then evaluated, which provides multi-dimensional considerations for the analysis (Wu et al., 2013). According to the feedback of learning analysis and personal learning habits, the system can automatically analyze and equip students with individual adaptable and personalized learning content.

3.3.3. Tools and Method Research in Learning Analytics

The tools and methods used in learning analytics are means for data analysis, which are the key process in learning analytics. The aim and method of learning analytics are both to extract usable information from the raw data in order to better understand both the learner's and the learning environment. The content analysis method is used to quantitatively analyze learners' characteristics from a sociological perspective; The social network analysis method is adopted to focus on the information distribution and learning progress in network learning and analyze the social structure from multiple perspectives; Through text mining and emotion analysis, students' learning status and learning attitudes are acquired, and their performances are evaluated (Li, Ma, & Huang, 2012). The use of visual analytics helps to convert chaotic data and then display the worked data visually in the interface through certain analytical reasoning techniques. Due to its strongly collaborative and interactive features, computer-supported collaborative learning (CSCL) is often used as cognitive tools and collaboration platform for learners in learning analytics period.

3.3.4. Technical Support Research in Learning Analytics

In the existing technology development fields, key big-data technologies and social analysis techniques are the two core analysis techniques. Open source tools for cloud computing, such as MapReduce and Hadoop, are important tools for learning analytics (Meng & Ci, 2015). In the process of data collection, mobile internet links to every person at every moment. And cloud storage technology provides a way out for technology acquisition and storage. The combination of cloud computing technology and cloud storage technology becomes important tools in learning analytics. In the integrated analysis system of learning analytics, which presented by international scholars, data analysis is achieved by studying dashboard and other visualization tools, as well as using multi-angle views to provide multi-angle information, in order to facilitate others to assess (Siemens et al., 2011). However, through the process of using internet for data acquisition and data storage, it inevitably involves user privacy. With the lack of complete system and technology for data security now, it is still a pivotal issue to keep personal information secret (Slade & Prinsloo, 2013).

3.3.5. Application Research in Learning Analytics

The basis of learning analysis application is its function to process big data and its aim of guiding learning. The international definition about Learning Analysis points out that what distinguishes Learning Analysis from the characteristics of other similar concepts is that it focuses on how to optimize on-line learning opportunities and how to make actual optimization towards learning processes and learning environment. It's concentration on procedure analysis makes it an effective teaching evaluation mode. This feature combines with characteristics of online courses such as the scale, openness and online characteristics represented by MOOCs and it can provide various evaluation models for teaching performance. The multi-perspective presentation of the results of learning analytics provides multi-angle data to build various evaluation modes for managers, tutors and researchers, which is representatively reflected in higher

education. What's more, the function of learning analysis to process big data is the basis of machine learning. The implement of learning analysis into machine learning will facilitate artificial intelligence to cater the needs of education.

3.4. Analysis of the Strategic Coordinate

The strategic coordinate takes the centrality and density as the horizontal and vertical axis respectively. It is used to show the strength of the association between different areas as well as the internal members. There are four quadrants in the strategic coordinate. The center of the whole strategy coordinate is composed of the average number of centrality and density. According to the above thought, the centrality and density of each class are obtained by calculating the average number of them of each class. According to the weighted average of all groups, the center of the strategic coordinate is (16.54, 15.14). As can be seen from Figure 6, the class group "application study" is located in the first quadrant of the strategic coordinates, whose density and centrality are both high, showing that the degree of association of the internal theme is also high, and has a strong connection to the external study. As can be seen that the research on the application of learning analysis has been the core in the field to scholars.

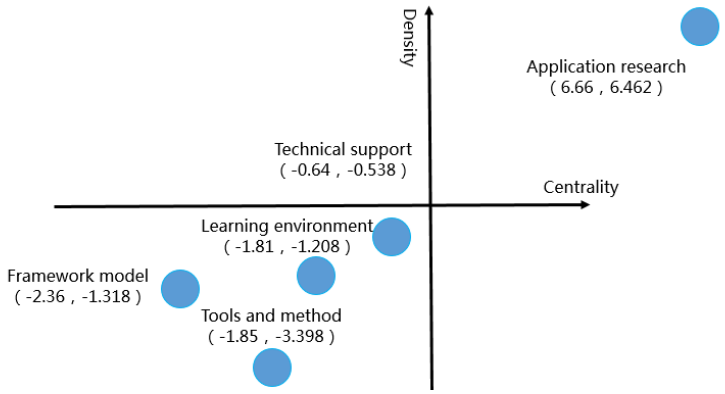

Figure 8. Strategic coordinate

Class groups "learning environment", "data analytics method", "technical support", "the framework model" all fell in the third quadrant. The density and the centrality are relatively low, which indicates that these studies are still not mature enough. They all fell on the edge of the network with non-significant impact to other study fields.

4. Conclusions and Implications

Through the theme analysis of researches on learning analysis above, it is clear that the application area of learning analysis technology is quite mature, while study on others are still insufficient. Researchers mainly delved into the application domain, with the expect to have an intensive study on the data of students' behaviors, the process of study and its happening context, and further to evaluate genuine functions of technology in educational areas. Looking at the outcomes of researches on learning analysis, we can see that there are still some deficiencies on learning analysis technology, with unbalance in different fields. Such research also has lots of challenges, including accuracy of data analysis, feedback mechanism, and the security of learning data.

Learning analysis technology is a method to obtain and analyze data on the basis of tools of computer or network, which also analyze students' behaviors using machines. However, the process of learning is complicated and changeable, and learning is a process of inner psychology and implicit thinking. That means learning can't be performed totally by explicit ways. Therefore, it seems that the behavior data can't reflect students' original appearance due to the absence of implicit data. The future research need to collect as much data as possible, and explain the meaning of data as extensive as possible to as well. In addition, what we should realize is that, as for learning analysis technology, different ways of explanation will have an impact on the accuracy of results of learning analysis.

The ethical issue of learning analysis technology is apparent, which may lead to complex privacy problems, and is

likely to be explained as wiretap. Since it investigates learners' learning behaviors secretly, learning analysis technology may cause invasion of privacy. Keeping watch on and collecting students' learning process and activities by camera not only has a negative influence on students' privacy, but makes students who are originally in state of nature feel uncomfortable and disturbed. Thus, the research in the future have to find the balance point between privacy and learning analysis technology, and integrate teaching and ethical issues during the process of learning analysis.

In conclusion, learning analysis technology has positive application values and great developmental potential. It is still on the initial stage of application facing a great amount of challenges. Objectively speaking, there is still a long way to go for learning analysis technology to realize its potential in assessment and improvement of learning,

Acknowledgements

We thank participating teachers and students. This research work is supported by National Natural Science Foundation of China (NSFC: 61075048), and Special Found for Beijing Common Construction Project, and the Fundamental Research Funds for the Central Universities (SKZZY2014094).

References

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining*, *1*(1), 3-17.

Ferguson, R. (2012). The state of learning analytics in 2012: A review and future challenges. *Knowledge Media Institute, Technical Report KMI-2012*, 1, 2012.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, *15*(3), 42-57.

Ifenthaler, D., & Widanapathirana, C. (2014). development and validation of a learning analytics framework: two case studies using support vector machines. *Technology, Knowledge and Learning*, *19*(1-2), 221-240.

Kim, M., & Lee, E. (2012). A multidimensional analysis tool for visualizing online interactions. *Journal of Educational Technology & Society*, *15*(3), 89-102.

Li, Y., Ma, S., & Huang, R. (2012). Learning analytics: Serving the learning process design and optimization. *Open Education Research*, *5*, 006.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588-599.

Meng, X. F., & Ci, X. (2013). Big data management: concepts, techniques and challenges. *Journal of Computer Research and Development*, *50*(1), 146-169.

Siemens, G. (2009). Technologically externalized knowledge and learning. *Connectives: Networked and social learning*.

Siemens, G., & d Baker, R. S. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252-254). ACM.

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., & Baker, R. S. J. D. (2011). Open learning analytics: an integrated & modularized platform. *Proposal to Design, Implement and Evaluate an Open Platform to Integrate Heterogeneous Learning Analytics Techniques.*

Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529.

Wu, Y., Chen, D., Ma, X., Cao, P., Feng, X., & Zhu, Z. (2013). Learning analytics: the new coming wave of education information. *Journal of Distance Education*, *31*(4), 11-19.