

REVIEW ARTICLE

Open Access

Big data in education: a state of the art, limitations, and future research directions



Maria Ijaz Baig, Liyana Shuib*  and Elaheh Yadegaridehkordi

* Correspondence: liyanashuib@gmail.com; liyanashuib@um.edu.my
Department of Information Systems,
Faculty of Computer Science &
Information Technology University
of Malaya, 50603 Kuala Lumpur,
Malaysia

Abstract

Big data is an essential aspect of innovation which has recently gained major attention from both academics and practitioners. Considering the importance of the education sector, the current tendency is moving towards examining the role of big data in this sector. So far, many studies have been conducted to comprehend the application of big data in different fields for various purposes. However, a comprehensive review is still lacking in big data in education. Thus, this study aims to conduct a systematic review on big data in education in order to explore the trends, classify the research themes, and highlight the limitations and provide possible future directions in the domain. Following a systematic review procedure, 40 primary studies published from 2014 to 2019 were utilized and related information extracted. The findings showed that there is an increase in the number of studies that address big data in education during the last 2 years. It has been found that the current studies covered four main research themes under big data in education, mainly, learner's behavior and performance, modelling and educational data warehouse, improvement in the educational system, and integration of big data into the curriculum. Most of the big data educational researches have focused on learner's behavior and performances. Moreover, this study highlights research limitations and portrays the future directions. This study provides a guideline for future studies and highlights new insights and directions for the successful utilization of big data in education.

Keywords: Data science applications in education, Learning communities, Teaching/learning strategies

Introduction

The world is changing rapidly due to the emergence of innovational technologies (Chae, 2019). Currently, a large number of technological devices are used by individuals (Shorfuzzaman, Hossain, Nazir, Muhammad, & Alamri, 2019). In every single moment, an enormous amount of data is produced through these devices (ur Rehman et al., 2019). In order to cater for this massive data, current technologies and applications are being developed. These technologies and applications are useful for data analysis and storage (Kalaian, Kasim, & Kasim, 2019). Now, big data has become a matter of interest for researchers (Anshari, Alas, & Yunus, 2019). Researchers are trying to define and characterize big data in different ways (Mikalef, Pappas, Krogstie, & Giannakos, 2018).

According to Yassine, Singh, Hossain, and Muhammad (2019), big data is a large volume of data. However, De Mauro, Greco, and Grimaldi (2016) referred to it as an informational asset that is characterized by high quantity, speed, and diversity. Moreover, Shahat (2019) described big data as large data sets that are difficult to process, control or examine in a traditional way. Big data is generally characterized into 3 Vs which are Volume, Variety, and Velocity (Xu & Duan, 2019). The volume refers to as a large amount of data or increasing scale of data. The size of big data can be measured in terabytes and petabytes (Herschel & Miori, 2017). In order to cater for the large volume of data, high capacity storage systems are required. The variety refers to as a type or heterogeneity of data. The data can be in a structured format (databases) or unstructured format (images, video, emails). Big data analytical tools are helpful in handling unstructured data. Velocity refers to as the speed at which big data can access. The data is virtually present in a real-time environment (Internet logs) (Sivarajah, Kamal, Irani, & Weerakkody, 2017).

Currently, the concept of 3 V's is inflated into several V's. For instance, Demchenko, Grosso, De Laat, and Membrey (2013) classified big data into 5vs, which are Volume, Velocity, Variety, Veracity, and Value. Similarly, Saggi and Jain (2018) characterized big data into 7 V's namely Volume, Velocity, Variety, Valence, Veracity, Variability, and Value.

Big data demand is significantly increasing in different fields of endeavour such as insurance and construction (Dresner Advisory Services, 2017), healthcare (Wang, Kung, & Byrd, 2018), telecommunication (Ahmed et al., 2018), and e-commerce (Wu & Lin, 2018). According to Dresner Advisory Services (2017), technology (14%), financial services (10%), consulting (9%), healthcare (9%), education (8%) and telecommunication (7%) are the most active sectors in producing a vast amount of data.

However, the educational sector is not an exception in this situation. In the educational realm, a large volume of data is produced through online courses, teaching and learning activities (Oi, Yamada, Okubo, Shimada, & Ogata, 2017). With the advent of big data, now teachers can access student's academic performance, learning patterns and provide instant feedback (Black & Wiliam, 2018). The timely and constructive feedback motivates and satisfies the students, which gives a positive impact on their performance (Zheng & Bender, 2019). Academic data can help teachers to analyze their teaching pedagogy and affect changes according to students' needs and requirement. Many online educational sites have been designed, and multiple courses based on individual student preferences have been introduced (Holland, 2019). The improvement in the educational sector depends upon acquisition and technology. The large-scale administrative data can play a tremendous role in managing various educational problems (Sorensen, 2018). Therefore, it is essential for professionals to understand the effectiveness of big data in education in order to minimize educational issues.

So far, several review studies have been conducted in the big data realm. Mikalef et al. (2018) conducted a systematic literature review study that focused on big data analytics capabilities in the firm. Mohammad & Torabi (2018), in their review study on big data, observed the emerging trends of big data in the oil and gas industry. Furthermore, another systematic literature review was conducted by Neilson, Daniel, and Tjandra (2019) on big data in the transportation system. Kamilaris, Kartakoullis, and Prenafeta-Boldú (2017), conducted a review study on the use of big data in agriculture. Similarly, Wolfert, Ge, Verdouw, and Bogaardt (2017) conducted a review study on the

use of big data in smart farming. Moreover, Camargo Fiorini, Seles, Jabbour, Mariano, and Sousa Jabbour (2018) conducted a review study on big data and management theory. Even though that many fields have been covered in the previous review studies, yet, a comprehensive review of big data in the education sector is still lacking today. Thus, this study aims to conduct a systematic review of big data in education in order to identify the primary studies, their trends & themes, as well as limitations and possible future directions. This research can play a significant role in the advancement of big data in the educational domain. The identified limitations and future directions will be helpful to the new researchers to bring encroachment in this particular realm.

The research questions of this study are stated below:

- 1) What are the trends in the papers published on big data in education?
- 2) What research themes have been addressed in big data in education domain?
- 3) What are the limitations and possible future directions?

The remainder of this study is organized as follows: Section 2 explains the review methodology and exposes the SLR results; Section 3 reports the findings of research questions; and finally, Section 4 presents the discussion and conclusion and research implications.

Review methodology

In order to achieve the aforementioned objective, this study employs a systematic literature review method. An effective review is based on analysis of literature, find the limitations and research gap in a particular area. A systematic review can be defined as a process of analyzing, accessing and understanding the method. It explains the relevant research questions and area of research. The essential purpose of conducting the systematic review is to explore and conceptualize the extant studies, identification of the themes, relations & gaps, and the description of the future directions accordingly. Thus, the identified reasons are matched with the aim of this study. This research applies the Kitchenham and Charters (2007) strategies. A systematic review comprised of three phases: Organizing the review, managing the review, and reporting the review. Each phase has specific activities. These activities are: 1) Develop review protocol 2) Formulate inclusion and exclusion criteria 3) Describe the search strategy process 4) Define the selection process 5) Perform the quality evaluation procedure and 6) Data extraction and synthesis. The description of each activity is provided in the following sections.

Review protocol

The review protocol provides the foundation and mechanism to undertake a systematic literature review. The essential purpose of the review protocol is to minimize the research bias. The review protocol comprised of background, research questions, search strategy, selection process, quality assessment, and extraction of data and synthesis. The review protocol helps to maintain the consistency of review and easy update at a later stage when new findings are incorporated. This is the most significant aspect that discriminates SLR from other literature reviews.

Inclusion and exclusion criteria

The aim of defining the inclusion and exclusion criteria is to be rest assured that only highly relevant researches are included in this study. This study considers the published articles in journals, workshops, conferences, and symposium. The articles that consist of introductions, tutorials and posters and summaries were eliminated. However, complete and full-length relevant studies published in the English language between January 2014 to 2019 March were considered for the study. The searched words should be present in title, abstract, or in the keywords section.

Table 1 shows a summary of the inclusion and exclusion criteria.

Search strategy process

The search strategy comprised of two stages, namely S1 (automatic stage) and S2 (manual stage). Initially, an automatic search (S1) process was applied to identify the primary studies of big data in education. The following databases and search engines were explored: Science Direct, SAGE.

Journals, Emerald Insight, Springer Link, IEEE Xplore, ACM Digital Library, Taylor and Francis and AIS e-Library. These databases were considered as it possessed highest impact journals and germane conference proceedings, workshops and symposium. According to Kitchenham and Charters (2007), electronic databases provide a broad perspective on a subject rather than a limited set of specific journals and conferences. In order to find the relevant articles, keywords on big data and education were searched to obtain reliable results. The general words correlated to education were also explored (education OR academic OR university OR learning.

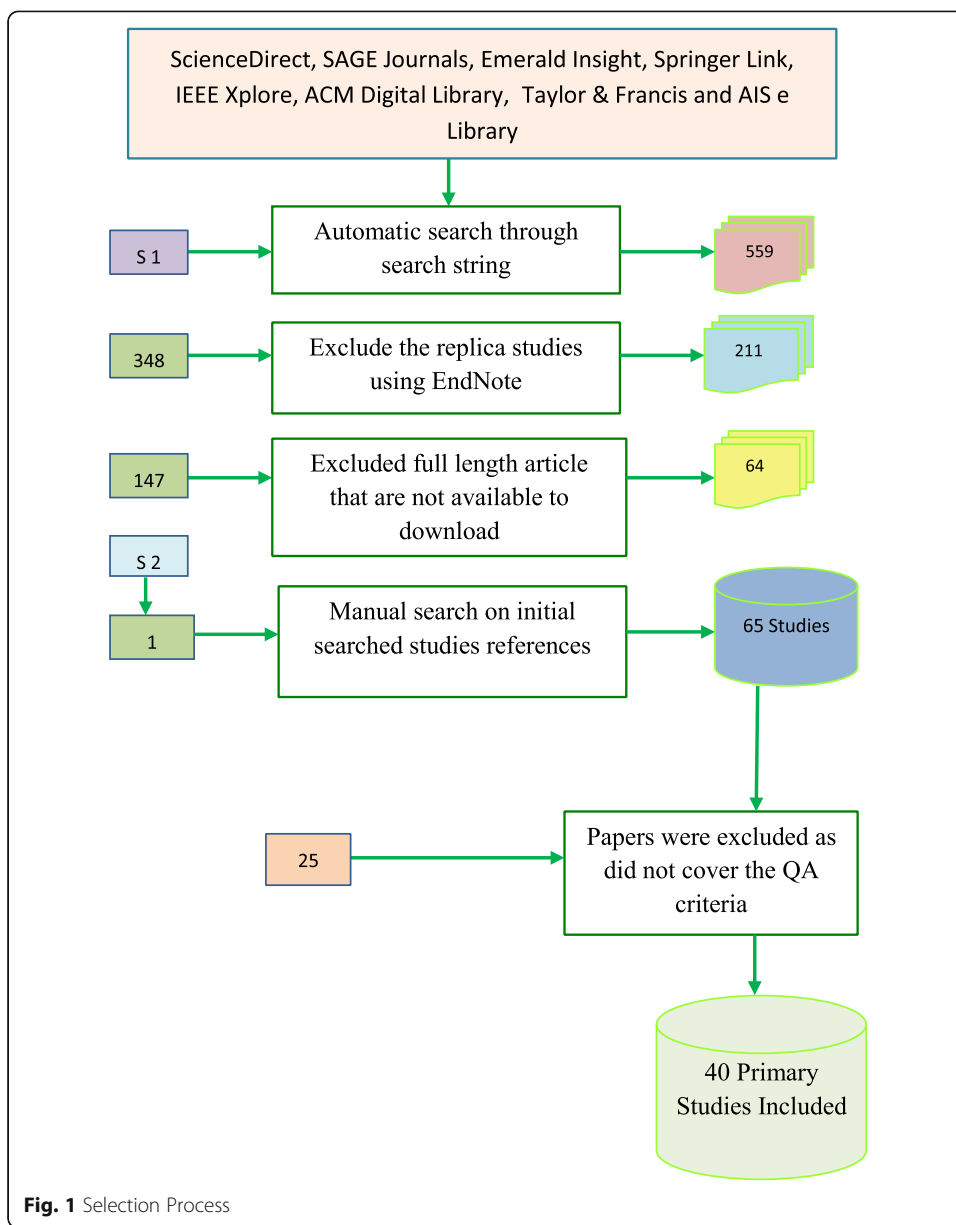
OR curriculum OR higher education OR school). This search string was paired with big data. The second stage is a manual search stage (S2). In this stage, a manual search was performed on the references of all initial searched studies. Kitchenham (2004) suggested that manual search should be applied to the primary study references. However, EndNote was used to manage, sort and remove the replicate studies easily.

Selection process

The selection process is used to identify the researches that are relevant to the research questions of this review study. The selection process of this study is presented in Fig. 1. By applying the string of keywords, a total number of 559 studies were found through automatic search. However, 348 studies are replica studies and were removed using the EndNote library. The inclusion and exclusion criteria were applied to the remaining 211 studies. According to Kitchenham and Charters (2007), recommendation and irrelevant studies should be excluded from the review subject. At this phase, 147 studies

Table 1 Inclusion and exclusion criteria

Inclusion	Exclusion
Complete and full-length studies	Incomplete and full length is not available to download
Published between January 2014 and March 2019	Published beyond this period
Published in the English language	Published other than English
Relevant and searched words present in title, abstracts or in the keywords section	Not relevant and searched words are not present in the title, abstracts or in the keywords section



were excluded as full-length articles were not available to download. Thus, 64 full-length articles were present to download and were downloaded. To ensure the comprehensiveness of the initial search results, the snowball technique was used. In the second stage, manual search (S2) was performed on the references of all the relevant papers through Google Scholar (Fig. 1). A total of 1 study was found through Google Scholar search. The quality assessment criteria were applied to 65 studies. However, 25 studies were excluded, as these studies did not fulfil the quality assessment criteria. Therefore, a total of 40 highly relevant primary studies were included in this research. The selection of studies from different databases and sources before and after results retrieval is shown in Table 2. It has been found that majority of research studies were present in Science Direct (90), SAGE Journals (50), Emerald Insight (81), Springer Link (38), IEEE

Xplore (158), ACM Digital Library (73), Taylor and Francis (17) and AIS e-Library (52). Google Scholar was employed only for the second round of manual search.

Quality assessment

According to (Kitchenham & Charters, 2007), quality assessment plays a significant role in order to check the quality of primary researches. The subtleties of assessment are totally dependent on the quality of the instruments. This assessment mechanism can be based on the checklist of components or a set of questions. The primary purpose of the checklist of components and a set of questions is to analyze the quality of every study. Nonetheless, for this study, four quality measurements standard was created to evaluate the quality of each research. The measurement standards are given as:

QA1. Does the topic address in the study related to big data in education?

QA2. Does the study describe the context?

QA3. Does the research method given in the paper?

QA4. Does data collection portray in the article?

The four quality assessment standards were applied to 65 selected studies to determine the integrity of each research. The measurement standards were categorized into low, medium and high. The quality of each study depends on the total number of score. Each quality assessment has two-point scores. If the study meets the full standard, a score of 2 is awarded. In the case of partial fulfillment, a score of 1 is acquired. If none of the assessment standards is met, then a score of 0 is awarded. In the total score, if the study gets below 4, it is counted as 'low' and exact 4 considered as 'medium'. However, the above 4 is reflected as 'high'. The details of studies are presented in Table 11 in [Appendix B](#). The 25 studies were excluded as it did not meet the quality assessment standard. Therefore, based on the quality assessment standard, a total of 40 primary studies were included in this systemic literature review (Table 10 in [Appendix A](#)). The scores of the studies (in terms of low, medium and high) are presented in [Fig. 2](#).

Table 2 Selection process results

Databases	Before results	After results
IEEE Xplore	158	8
ScienceDirect	90	14
Emerald insight	81	4
AIS Electronic Library	52	1
Sage	50	5
ACM Digital Library	73	5
Springer Link	38	1
Taylor and Francis	17	2
Google Scholar	1	–
		40

Data extraction and synthesis

The data extraction and synthesis process were carried by reading the 65 primary studies. The studies were thoroughly studied, and the required details extracted accordingly. The objective of this stage is to find out the needed facts and figure from primary studies. The data was collected through the aspects of research ID, names of author, the title of the research, its publishing year and place, research themes, research context, research method, and data collection method. Data were extracted from 65 studies by using this aspect. The narration of each item is given in Table 3. The data extracted from all primary studies are tabulated. The process of data synthesizing is presented in the next section.

Findings

What are the trends in the papers published on big data in education?

Figure 3 presented the allocation of studies based on their publication sources. All publications were from high impact journals, high-level conferences, and workshops. The primary studies are comprised of 21 journals, 17 conferences, 1 workshop, and 1 symposium. However, 14 studies were from Science Direct journals and conferences. A total of 5 primary studies were from the SAGE group, 1 primary study from Springer-Link. Whereas 6 studies were from IEEE conferences, 2 studies were from IEEE symposium and workshop. Moreover, 1 primary study from AISel Conference. Hence, 4 studies were from Emeraldinsight journals, 5 studies were from ACM conferences and 2 studies were from Taylor and Francis. The summary of published sources is given in Table 4.

Temporal view of researches

The selection period of this study is from January 2014–March 2019. The yearly allocation of primary studies is presented in Fig. 4. The big data in education trend started in the year 2014. This trend gradually gained popularity. In 2015, 8 studies were published in this domain. It has been found that a number of studies rise in the year 2017. Thus, the highest number of publication in big data in the education realm was observed in the year 2017. In 2017, 12 studies were published. This trend continued in 2018, and in that year, 11 studies that belong to big data in education were published. In 2019, the trend of this domain is still continued as this paper covers that period of March 2019. Thus, 4 studies were published until March 2019.

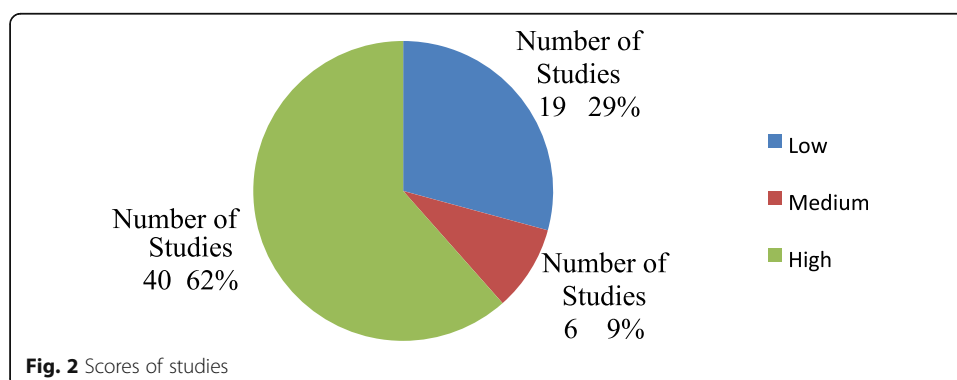


Table 3 The narration of data extraction Items

Data Extraction Items	Narration
Research ID	To give a distinctive number to research paper
Author Names	The creator of the study
Study Title	The name of the study
Publication Year	The year of publication (e.g. 2019)
Publication Place	E.g. Journal, Conference, and workshop, etc.
Research Theme	The particular domain/area of study (e.g. e-learning, student grading systems, etc)
Research Context	The theoretical background, perspective or framework
Research Method	Quantitative, qualitative and mix method, etc.
Data Collection Method	Survey, interviews and observations, etc.

Citation

In order to find the total citation count for the studies, Google Scholar was used. The number of citation is shown in Fig. 5. It has been observed that 28 studies were cited by other sources 1–50 times. However, 11 studies were not cited by any other source. Thus, 1 study was cited by other sources 127 times. The top cited studies with their titles are presented in Table 5, which provides general verification. The data provided here is not for comparison purpose among the studies.

Research methodologies

The research methods employed by primary studies are shown in Fig. 6. It has been found that majority of them are review based studies. These reviews were conducted in a different educational context and big data. However, reviews covered 28% of primary studies. The second most used research method was quantitative. This method covered 23% of the total primary studies. Only 3% of the study was based on a mix method approach. Moreover, design science method also covered 3% of primary studies. Nevertheless, 20% of the studies used qualitative research method, whereas the remaining 25% of the studies were not discussed and given in the articles.

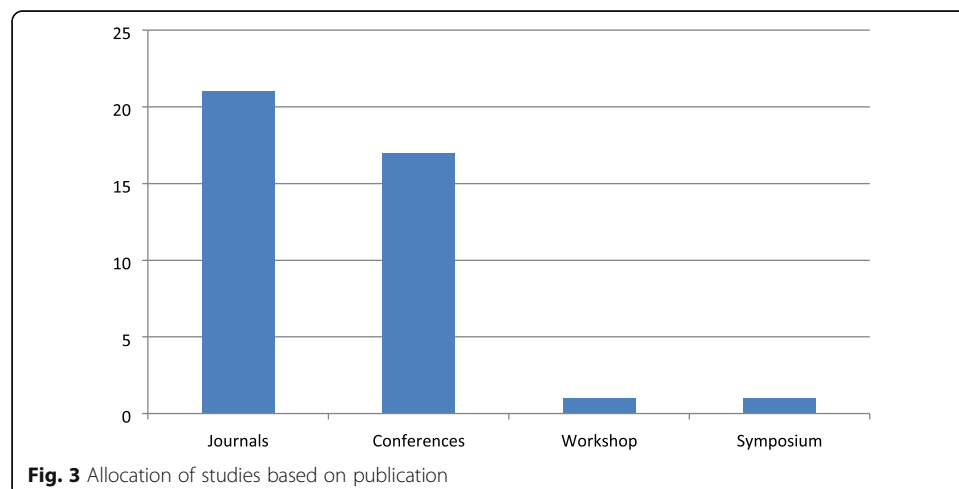
**Fig. 3** Allocation of studies based on publication

Table 4 Summary of publication sources

Publications	Sources
9	Science Direct journals
5	SAGE Journals
1	SpringerLink Journals
6	IEEE conferences
5	ACM conferences
1	IEEE Symposium
4	Emeraldinsight journals
5	Science Direct Conferences
1	IEEE Workshop
1	AISEL Conference
2	Taylor and Francis Journals

Data collection methods

The data collection methods used by primary studies are shown in Fig. 7. The primary studies employed different data collection methods. However, the majority of studies used extant literature. The 5 types of research conducted surveys which covered 13% of primary Studies. The 4 studies carried experiments for data collection, which covered 10% of primary studies. Nevertheless, 6 studies conducted interviews for data collection, which is based on 15% of primary studies. The 4 studies used data logs which are based on 10% of primary studies. The 2 studies collected data through observations, 1 study used social network data, and 3 studies used website data. The observational, social network data and website-based researches covered 5%, 3% and 8% of primary studies. Moreover, 11 studies used extant literature and 1 study extracted data from a focus group discussion. The extant literature and focus group-based studies covered 28% and 3% of primary studies. However, the data collection method is not available for the remaining 3 studies.

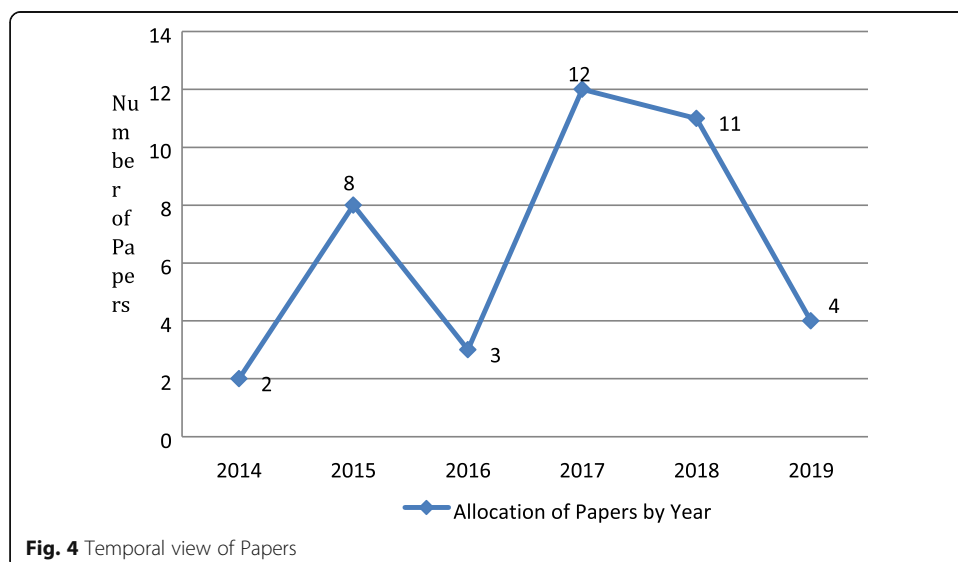
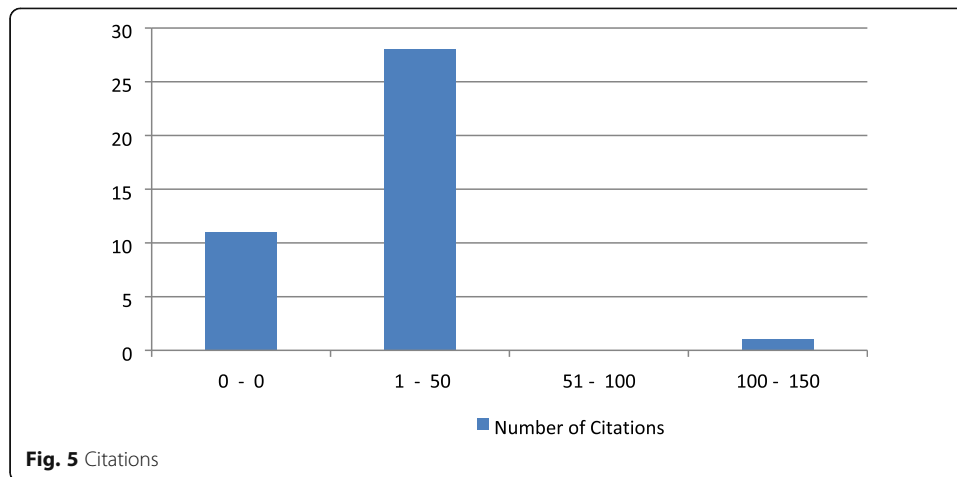


Fig. 4 Temporal view of Papers



What research themes have been addressed in educational studies of big data?

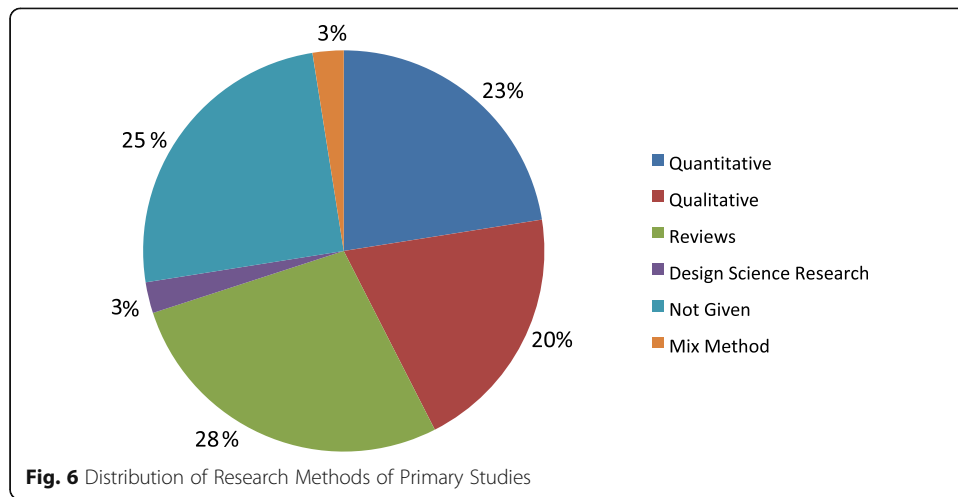
The theme refers to an idea, topic or an area covered by different research studies. The central idea reflects the theme that can be helpful in developing real insight and analysis. A theme can be in single or combination of more words (Rimmon-Kenan, 1995). This study classified big data research themes into four groups (Table 6). Thus, Fig. 8 shows a mind map of big data in education research themes, sub-themes, and the methodologies.

Figure 9 presents, research themes under big data in education, namely learner's behavior and performance, modelling, and educational data warehouse, improvement of the educational system, and integration of big data into the curriculum.

The first research theme was based on the learner's behavior and performance. This theme covers 21 studies, which consists of 53% of overall primary studies (Fig. 9). The theme studies are based on teaching and learning analytics, big data frameworks, user behaviour, and attitude, learner's strategies, adaptive learning, and satisfaction. The total number of 8 studies relies on teaching and learning analytics (Table 7). Three (3)

Table 5 Top cited studies

R-ID	Title	Citation
R4	The role of big data and cognitive computing in the learning process	23
R8	Using Big Data in the Academic Environment	22
R32	Internet use that reproduces educational inequalities: Evidence from big data	15
R40	Big Data Application in Education: Dropout Prediction in Edx MOOCs	28
R36	Discovering Big Data Modeling for Educational World	25
R21	The Life Between Big Data Log Events	23
R33	Mining theory-based patterns from Big data: Identifying self regulated learning strategies in Massive Open Online Courses	53
R23	CS principles goes to middle school	29
R5	Toward an integration of Big Data, technology and information systems competencies into the accounting curriculum	54
R27	Education and training for a successful career in Big Data and Business Analytics	51
R38	Data entry: towards the critical study of digital data and education	168



studies deal with big data framework. However, 6 studies concentrated on user behaviour and attitude. Nevertheless, 2 studies dwell on learning strategies. The adaptive learning and satisfaction covered 1 study, respectively. In this theme, 2 studies conducted surveys, 4 studies carried out experiments and 1 study employed the observational method. The 5 studies reported extant literature. In addition, 4 studies used event log data and 5 conducted interviews (Fig. 10).

In the second theme, studies conducted focused on modeling and educational data warehouses. In this theme, 6 studies covered 15% of primary studies. This theme studies investigated the cloud environment, big data modeling, cluster analysis, and data warehouse for educational purpose (Table 8). Three (3) studies introduced big data modeling in education and highlighted the potential for organizing data from multiple sources. However, 1 study analyzed data warehouse with big data tools (Hadoop). Moreover, 1 study analyzed the accessibility of huge academic data in a cloud computing environment whereas, 1 study used clustering techniques and data warehouse for educational purpose. In this theme, 4 studies reported extant review, 1 study conduct survey, and 1 study used social network data.

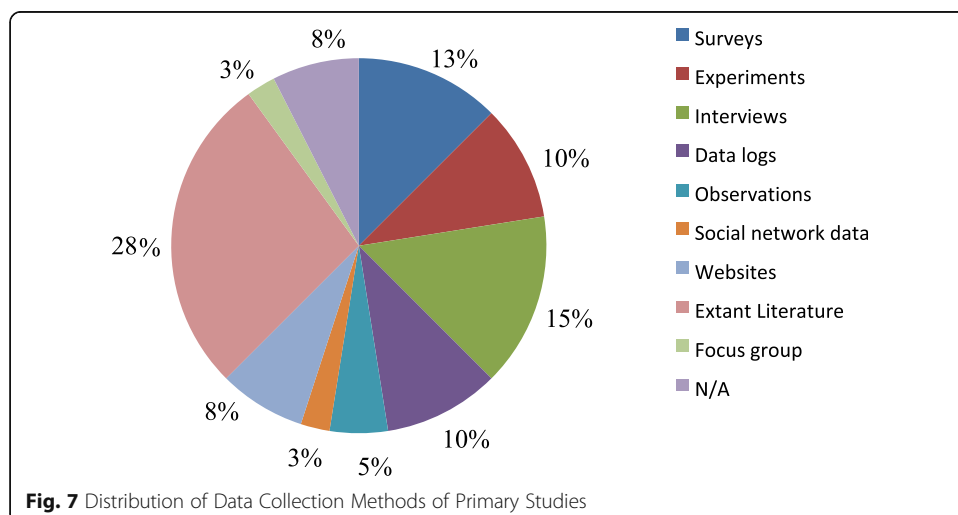


Table 6 Description of Research Themes

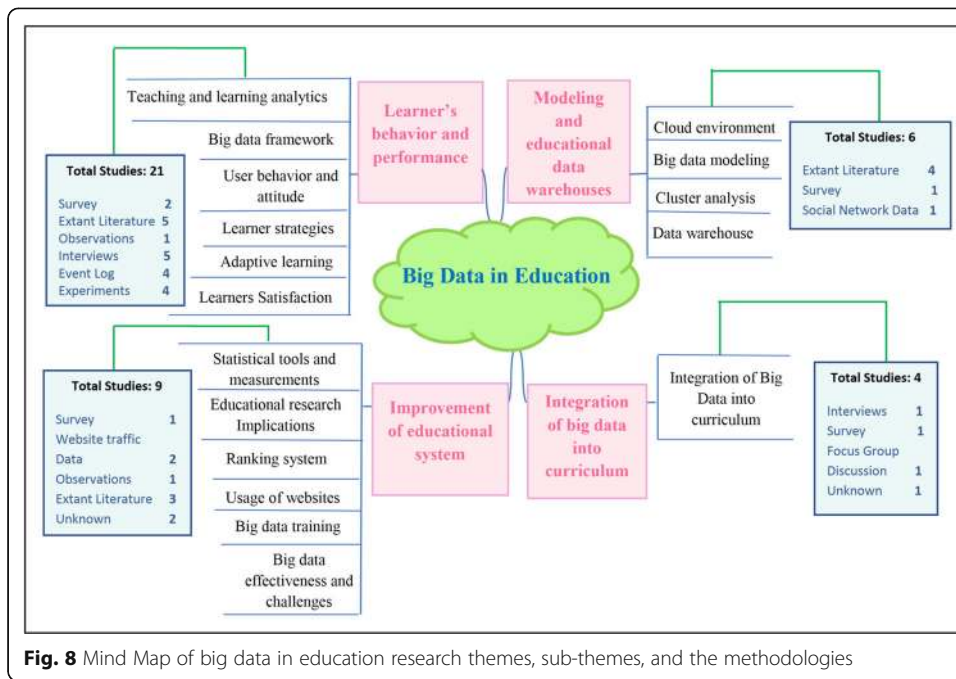
Studies Themes	Description	References
Learners behavior and performance	Studies that investigated the learner's attitude, satisfaction, strategies, behavior, big data frameworks, adaptive learning, teaching and learning analytics.	(Cantabella, Martínez-España, Ayuso, Yáñez, & Muñoz, 2019; Chaurasia & Frieda Rosin, 2017; Chaurasia, Kodwani, Lachhwani, & Ketkar, 2018; Coccoli, Maresca, & Stanganelli, 2017; Dessì, Fenu, Marras, & Reforgiato Recupero, 2019; Dubey & Gunasekaran, 2015; Elia, Solazzo, Lorenzo, & Passiante, 2018; Hirashima, Supianto, & Hayashi, 2017; Lia & Zhaia, 2018; Liang, Yang, Wu, Li, & Zheng, 2016; Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-Gama, 2018; Muthukrishnan & Yasin, 2018; Nie et al., 2018; Oi et al., 2017; Roy & Singh, 2017; Sedkaoui & Khelfaoui, 2019; Shorfuzzaman et al., 2019; Su, Ding, Lue, Lai, & Su, 2017; Troisi, Grimaldi, Loia, & Maione, 2018; Veletsianos, Reich, & Pasquini, 2016; Yang & Du, 2016)
Modeling and educational Data Warehouse	Studies that introduced big data modeling and analyzed big data tools (Hadoop) with data warehouses. It explored the cloud environment and cluster analysis for accessibility and processing of educational data.	(Logica & Magdalena, 2015; Pardos, 2017; Petrova-Antonova, Georgieva, & Ilieva, 2017; Ramos, Machado, & Cordeiro, 2015; Santoso & Yulia, 2017; Wassan, 2015)
Improvement of the educational system	Studies that analyzed statistical tools, measurements, challenges, and ICT effectiveness. It emphasized on training and various implications. It introduced a ranking system and observes the usage of websites to improve the educational system.	(Gupta & Rani, 2018; Martínez-Abad, Gamazo, & Rodríguez-Conde, 2018; Ong, 2015; Ozgur, Kleckner, & Li, 2015; Qiu, Huang, & Patel, 2015; Selwyn, 2014; Sooriamurthi, 2018; Sorensen, 2018; Zhang, 2015)
Integration of big data into the curriculum	Studies that introduce big data topics into different courses and highlighted its implications for education.	(Buffum et al., 2014; Dinter, Jaekel, Kollwitz, & Wache, 2017; Nelson & Pouchard, 2017; Sledgianowski, Gomaa, & Tan, 2017)

The third theme concentrated on the improvement of the educational system. In this theme, 9 studies covered 23% of the primary studies. They consist of statistical tools and measurements, educational research implications, big data training, the introduction of the ranking system, usage of websites, big data educational challenges and effectiveness (Table 9). Two (2) studies considered statistical tools and measurements. Educational research implications, ranking system, usage of websites, and big data training covered 1 study respectively. However, 3 studies considered big data effectiveness and challenges. In this theme, 1 study conducted a survey for data collection, 2 studies used website traffic data, and 1 study exploited the observational method. However, 3 studies reported extant literature.

The fourth theme concentrated on incorporating the big data approaches into the curriculum. In this theme, 4 studies covered 10% of the primary studies. These 4 studies considered the introduction of big data topics into different courses. However, 1 study conducted interviews, 1 study employed survey method and 1 study used focus group discussion.

What are the limitations and possible future directions?

The 20% of the studies (Fig. 6) used qualitative research methods (Dinter et al., 2017; Veletsianos et al., 2016; Yang & Du, 2016). Qualitative methods are mostly applicable to observe the single variable and its relationship with other variables. However, this



method does not quantify relationships. In qualitative researches, understanding is attained through 'wording' (Chaurasia & Frieda Rosin, 2017). The behaviors, attitude, satisfaction, performance, and overall learning performance are related with human phenomena (Cantabella et al., 2019; Elia et al., 2018; Sedkaoui & Khelifaoui, 2019). Qualitative researches are not statistically tested (Chaurasia & Frieda Rosin, 2017). Big data educational studies which employed qualitative methods lacks some certainties that are present in quantitative research methods. Therefore, future researches might quantify the educational big data applications and its impact on higher education.

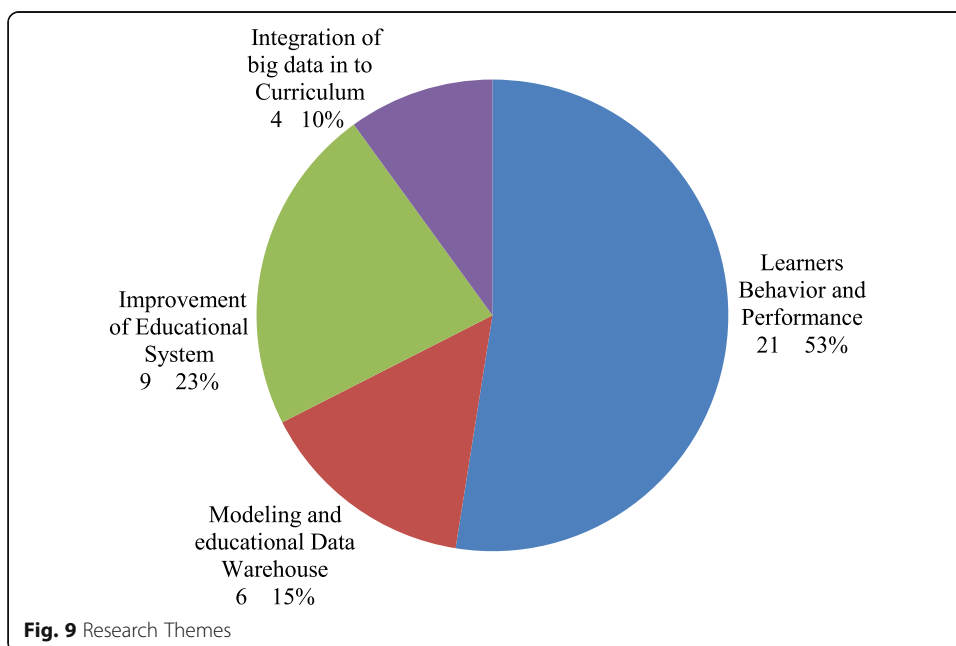


Table 7 Learner’s behavior and performance studies

Learners behavior and performance	Number of Studies
Teaching and learning analytics	8
Big data framework	3
User behavior and attitude	6
Learner strategies	2
Adaptive learning	1
Learners satisfaction	1

The six studies conducted interviews for data collection (Chaurasia et al., 2018; Chaurasia & Frieda Rosin, 2017; Nelson & Pouchard, 2017; Troisi et al., 2018; Veletsianos et al., 2016). However, 2 studies used observational method (Maldonado-Mahauad et al., 2018; Sooriamurthi, 2018) and one (1) study conducted focus group discussion (Buffum et al., 2014) for data collection (Fig. 10). The observational studies were conducted in uncontrolled environments. Sometimes results of these studies lead to self-selection biased. There is a chance of ambiguities in data collection where human language and observation are involved. The findings of interviews, observations and focus group discussions are limited and cannot be extended to a wider population of learners (Dinter et al., 2017).

The four big data educational studies analyzed the event log data and conducted interviews (Cantabella et al., 2019; Hirashima et al., 2017; Liang et al., 2016; Yang & Du, 2016). However, longitudinal data are more appropriate for multidimensional measurements and to analyze the large data sets in the future (Sorensen, 2018).

The eight studies considered the teaching and learning analytics (Chaurasia et al., 2018; Chaurasia & Frieda Rosin, 2017; Dessì et al., 2019; Roy & Singh, 2017). There are limited researches that covered the aspects of learning environments, ethical and cultural values and government support in the adoption of educational big data (Yang & Du, 2016). In the future, comparison of big data in different learning environments, ethical and cultural values, government support and training in adopting big data in higher education can be covered through leading journals and conferences.

The three studies are related to big data frameworks for education (Cantabella et al., 2019; Muthukrishnan & Yasin, 2018). However, the existed frameworks did not cover the organizational and institutional cultures, yet lacking robust theoretical grounds

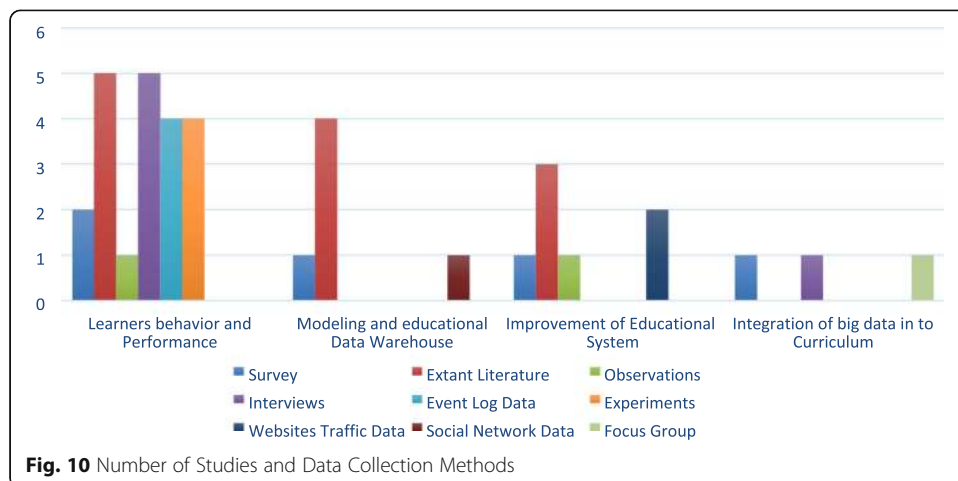


Fig. 10 Number of Studies and Data Collection Methods

Table 8 Modeling and educational data warehouse studies

Modeling and educational data warehouses	Number of Studies
Cloud Environment	1
Big data Modeling	3
Cluster Analysis	1
Data warehouse	1

(Dubey & Gunasekaran, 2015; Muthukrishnan & Yasin, 2018). In the future, big data educational framework that concentrates on theories and adoption of big data technology is recommended. The extension of existed models and interpretation of data models are recommended. This will help in better decision and ensure the predictive analysis in the academic realm. Moreover, further relations can be tested by integrating other constructs like university size and type (Chaurasia et al., 2018).

The three studies dwelled on big data modeling (Pardos, 2017; Petrova-Antonova et al., 2017; Wassan, 2015). These models do not incorporate with the present systems (Santoso & Yulia, 2017). Therefore, efficient research solutions that can manage the educational data, new interchanging and resources are required in the future. One (1) study explored a cloud-based solution for managing academic big data (Logica & Magdalena, 2015). However, this solution is expensive. In the future, a combination of LMS that is supported by open-source applications and software's can be used. This development will help universities to obtain benefits from unified LMS and to introduce new trends and economic opportunities for the academic industry. The data warehouse with big data tools was investigated by one (1) study (Santoso & Yulia, 2017). Nevertheless, a manifold node cluster can be implemented to process and access the structural and un-structural data in future (Ramos et al., 2015). In addition, new techniques that are based on relational and nonrelational databases and development of index catalogs are recommended to improve the overall retrieval system. Furthermore, the applicability of the least analytical tools and parallel programming models are needed to be tested for academic big data. MapReduce, MongoDB, pig,

Cassandra, Yarn, and Mahout are suggested for exploring and analysis of educational big data (Wassan, 2015). These tools will improve the analysis process and help in the development of reliable models for academic analytics.

One (1) study detected ICT factors through data mining techniques and tools in order to enhance educational effectiveness and improves its system (Martínez-Abad et al., 2018). Additionally, two studies also employed big data analytic tools on popular websites to examine the academic user's interest (Martínez-Abad et al., 2018; Qiu et al., 2015). Thus, in future research, more targeted strategies and regions can be selected

Table 9 Improvement of educational system theme studies

Improvement of the educational system	Number of Studies
Statistical tools and measurements	2
Educational research Implications	1
Ranking system	1
Usage of websites	1
Big data training	1
Big data effectiveness and challenges	3

for organizing the academic data. Similarly, in-depth data mining techniques can be applied according to the nature of the data. Thus, the foreseen research can be used to validate the findings by applying it on other educational websites. The present research can be extended by analyzing the socioeconomic backgrounds and use of other websites (Qiu et al., 2015).

The two research studies were conducted on measurements and selection of statistical software for educational big data (Ozgur et al., 2015; Selwyn, 2014). However, there is no statistical software that is fit for every academic project. Therefore, in future research, all in one' type statistical software is recommended for big data in order to fulfill the need of all academic projects. The four research studies were based on incorporating the big data academic curricula (Buffum et al., 2014; Sledgianowski et al., 2017). However, in order to integrate the big data into the curriculum, the significant changes are required. Firstly, in future researches, curricula need to be redeveloped or restructured according to the level and learning environment (Nelson & Pouchard, 2017). Secondly, the training factor, learning objectives, and outcomes should be well designed in future studies. Lastly, comparable exercises, learning activities and assessment plan need to be well structured before integrating big data into curricula (Dinter et al., 2017).

Discussion and conclusion

Big data has become an essential part of the educational realm. This study presented a systematic review of the literature on big data in the educational sector. However, three research questions were formulated to present big data educational studies trends, themes, and identification of the limitations and directions for further research. The primary studies were collected by performing a systematic search through IEEE Xplore, ScienceDirect, Emerald Insight, AIS Electronic Library, Sage, ACM Digital Library, Springer Link, Taylor and Francis, and Google Scholar databases. Finally, 40 studies were selected that meet the research protocols. These studies were published between the years 2014 (January) and 2019 (April). Through the findings of this study, it can be concluded that 53% of extant studies were conducted on learner's behavior and performance theme. Moreover, 15% of the studies were on modeling and educational Data Warehouse, and 23% of the studies were on the improvement of educational system themes. However, only 10% of the studies were on the integration of big data into the curriculum theme.

Thus, a large number of studies were conducted in learner's behavior and performance theme. However, other themes gained lesser attention. Therefore, more researches are expected in modeling and educational Data Warehouse in the future, in order to improve the educational system and integration of big data into the curriculum, related themes.

It has been found that 20% of the studies used qualitative research methods. However, 6 studies conducted interviews, 2 studies used observational method and 1 study conducted focus group discussion for data collection. The findings of interviews, observations and focus group discussions are limited and cannot be extended to a wider population of learners. Therefore, prospect researches might quantify the educational big data applications and its impact in higher education. The longitudinal data are more appropriate for multidimensional measurements and future analysis of the large

data sets. The eight studies were carried out on teaching and learning analytics. In the future, comparison of big data in different learning environments, ethical and cultural values, government support and training to adopt big data in higher education can be covered through leading journals and conferences.

The three studies were related to big data frameworks for education. In the future, big data educational framework that dwells on theories and extension of existed models are recommended. The three studies concentrated on big data modeling. These models cannot incorporate with present systems. Therefore, efficient research solutions are that can manage the educational data, new interchanging and resources are required in a future study. The two studies explored a cloud-based solution for managing academic big data and investigated data warehouse with big data tools. Nevertheless, in the future, a manifold node cluster can be implemented for processing and accessing of the structural and un-structural data. The applicability of the least analytical tools and parallel programming models needs to be tested for academic big data.

One (1) study considered the detection of ICT factors through data mining technique and 2 studies employed big data analytic tools on popular websites to examine the academic user's interest. Thus, more targeted strategies and regions can be selected for organizing the academic data in future. Four (4) research studies featured on incorporating the big data academic curricula. However, the big data based curricula need to be redeveloped by considering the learning objectives. In the future, well-designed learning activities for big data curricula are suggested.

Research implications

This study has two folded implications for stakeholders and researchers. Firstly, this review explored the trends published on big data in education realm. The identified trends uncover the studies allocation, publication sources, sequential view and most cited papers. In addition, it highlights the research methods used in these studies. The described trends can provide opportunities and new ideas to researchers to predict the accurate direction in future studies.

Secondly, this research explored the themes, sub-themes, and the methodologies in big data in education domain. The classified themes, sub-themes, and the methodologies present a comprehensive overview of existing literature of big data in education. The described themes and sub-themes can be helpful for researchers to identify new research gap and avoid using repeated themes in future studies. Meanwhile, it can help researchers to focus on the combination of different themes in order to uncover new insights on how big data can improve the learning and teaching process. In addition, illustrated methodologies can be useful for researchers in the selection of method according to nature of the study in future.

Identified research can be an implication for stakeholders towards the holistic expansion of educational competencies. The identified themes give new insight to universities to plan mixed learning programs that combine conventional learning with web-based learning. This permits students to accomplish focused learning outcomes, engrossing exercises at an ideal pace. It can be helpful for teachers to apprehend the ways to gauge students learning behaviour and attitude simultaneously and advance teaching strategy accordingly. Understanding the latest trends in big data and education are of growing

importance for the ministry of education as they can develop flexible possibly to support the institutions to improve the educational system.

Lastly, the identified limitations and possible future directions can provide guidelines for researchers about what has been explored or need to explore in future. In addition, stakeholders can also extract ideas to impart the future cohort and comprehend the learning and academic requirements.

Appendix A

Table 10 Primary studies references

R-ID	References
R1	Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J. A., & Muñoz, A. (2019). Analysis of student behavior in learning management systems through a Big Data framework. <i>Future Generation Computer Systems</i> , 90 (2), 262–272.: https://doi.org/10.1016/j.future.2018.08.003
R2	Lia, Y., & Zhaia, X. (2018). Review and Prospect of Modern Education using Big Data. <i>Procedia Computer Science</i> , 129 (3), 341–347.: https://doi.org/10.1016/j.procs.2018.03.085
R3	Elia, G., Solazzo, G., Lorenzo, G., & Passiante, G. (2018). Assessing learners' satisfaction in collaborative online courses through a big data approach. <i>Computers in Human Behavior</i> . 92, 589–599.: https://doi.org/10.1016/j.chb.2018.04.033
R4	Coccoli, M., Maresca, P., & Stanganelli, L. (2017). The role of big data and cognitive computing in the learning process. <i>Journal of Visual Languages & Computing</i> , 38, 97–103.: https://doi.org/10.1016/j.jvlc.2016.03.002
R5	Sledgianowski, D., Gomaa, M., & Tan, C. (2017). Toward integration of Big Data, technology and information systems competencies into the accounting curriculum. <i>Journal of Accounting Education</i> , 38 (1), 8193.: https://doi.org/10.1016/j.jaccedu.2016.12.008
R6	Leo Willyanto Santoso, & Yulia. (2017). Data Warehouse with Big Data Technology for Higher Education. <i>Procedia Computer Science</i> , 124 (1), 93–99.: https://doi.org/10.1016/j.procs.2017.12.134
R7	Ramos, T. G., Machado, J. C. F., & Cordeiro, B. P. V. (2015). Primary Education Evaluation in Brazil Using Big Data and Cluster Analysis. <i>Procedia Computer Science</i> , 55 (1), 1031–1039.: https://doi.org/10.1016/j.procs.2015.07.061
R8	Logica, B., & Magdalena, R. (2015). Using Big Data in the Academic Environment. <i>Procedia Economics and Finance</i> , 33 (2), 277–286.: https://doi.org/10.1016/s2212-5671(15)01712-8
R9	Qiu, R. G., Huang, Z., & Patel, I. C. (2015, June). A big data approach to assessing the US higher education service. In <i>2015 12th International Conference on Service Systems and Service Management (ICSSSM)</i> (pp. 1–6). New York: IEEE. https://doi.org/10.1109/ICSSSM.2015.7170149
R10	Nelson, M., & Pouchard, L. (2017). A pilot “big data” education modular curriculum for engineering graduate education: Development and implementation. Paper presented at the Frontiers in Education Conference (FIE), Indianapolis, USA (pp. 1–5). United States: IEEE. https://doi.org/10.1109/FIE.2017.8190688
R11	Hirashima, T., Supianto, A. A., & Hayashi, Y. (2017, September). Modelbased approach for educational big data analysis of learners thinking with process data. In <i>2017 International Workshop on Big Data and Information Security (IW BIS)</i> (pp. 11–16). San Diego: IEEE. https://doi.org/10.1177/0165551518789880
R12	Roy, S., & Singh, S. N. (2017). Emerging trends in applications of big data in data mining and learning analytics. In <i>2017 7th Conference Cloud Computing, Data Science & Engineering-Confluence</i> (pp. -198). New York: IEEE. https://doi.org/10.1109/confluence.2017.7943148
R13	Ong, V. K. (2015). <i>Big Data and Its Research Implications for Higher Education: Cases from UK Higher Education Institutions</i> . Paper presented at the 2015 IIAI 4th International Confress on Advanced Applied Informatics (pp. 487–491). IEEE.: https://doi.org/10.1109/IIAI-AAI.2015.178
R14	Su, Y. S., Ding, T. J., Lue, J. H., Lai, C. F., & Su, C. N. (2017, May). Applying big data analysis technique to students' learning behavior and learning resource recommendation in a MOOCs course. In <i>2017 International conference on applied system innovation (ICASI)</i> (pp. 1229–1230). IEEE.: https://doi.org/10.1109/ICASI.2017.7988114
R15	Muthukrishnan, S. M., & Yasin, N. B. M. (2018). <i>Big Data Framework for Students' Academic</i> . Paper presented at the Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia (pp. 376–382). USA: IEEE. https://doi.org/10.1109/ISCAIE.2018.8405502
R16	Ozgur, C., Kleckner, M., & Li, Y. (2015). Selection of Statistical Software for Solving Big Data Problems. <i>SAGE</i>

Table 10 Primary studies references (Continued)

R-ID	References
	<i>Open</i> , 5 (2), 59–94.: https://doi.org/10.1177/2158244015584379
R17	Sorensen, L. C. (2018). "Big Data" in Educational Administration: An Application for Predicting School Dropout Risk. <i>Educational Administration Quarterly</i> , 45 (1), 1–93: https://doi.org/10.1177/0013161x18799439
R18	Yang, F., & Du, Y. R. (2016). Storytelling in the Age of Big Data. <i>Asia Pacific Media</i> , 26 (2), 148–162.: https://doi.org/10.1177/1326365x16673168
R19	Nie, M., Yang, L., Sun, J., Su, H., Xia, H., Lian, D., & Yan, K. (2018). Advanced forecasting of career choices for college students based on campus big data. <i>Frontiers of Computer Science</i> , 12 (3), 494–503.: https://doi.org/10.1007/s11704-017-6498-6
R20	Gupta, D., & Rani, R. (2018). A study of big data evolution and research challenges. <i>Journal of Information Science</i> . 45 (3), 322–340.: https://doi.org/10.1177/0165551518789880
R21	Veletsianos, G., Reich, J., & Pasquini, L. A. (2016). The Life Between Big Data Log Events. <i>AERA Open</i> , 2 (3), 1–45.: https://doi.org/10.1177/2332858416657002
R22	Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M. J. (2018). <i>Big Data in Education</i> . Paper presented at the Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM'18, Salamanca, Spain (pp. 145–150). New York: ACM. https://doi.org/10.1145/3284179.3284206
R23	Buffum, P. S., Martinez-Arocho, A. G., Frankosky, M. H., Rodriguez, F. J., Wiebe, E. N., & Boyer, K. E. (2014, March). CS principles goes to middle school: learning how to teach Big Data. In <i>Proceedings of the 45th ACM technical Computer science education</i> (pp. 151–156). ACM.: https://doi.org/10.1145/2538862.2538949
R24	Dinter, B., Jaekel, T., Kollwitz, C., & Wache, H. (2017). <i>Teaching Big Data Management – An Active Learning Approach for Higher Education</i> . Paper presented at the Proceedings of the Pre-ICIS 2017 SIGDSA. (pp. 1–17). AISEL.
R25	Chaurasia, S. S., & Frieda Rosin, A. (2017). From Big Data to Big Impact: analytics for teaching and learning in higher education. <i>Industrial and Commercial Training</i> , 49 (7), 321–328.: https://doi.org/10.1108/ict-10-2016-0069
R26	Chaurasia, S. S., Kodwani, D., Lachhwani, H., & Ketkar, M. A. (2018). Big academic and learning analytics. <i>International Journal of Management</i> , 32 (6), 1099–1117.: https://doi.org/10.1108/ijem-08-2017-0199
R27	Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in Big Data and Business Analytics. <i>Industrial and Commercial Training</i> , 47 (4), 174–181.: https://doi.org/10.1108/ict-08-2014-0059
R28	Sedkaoui, S., & Khelfaoui, M. (2019). Understand, develop and enhance the learning process with big data. <i>Information Discovery and Delivery</i> , 47 (1), 2–16.: https://doi.org/10.1108/idd-09-2018-0043
R29	Sooriamurthi, R. (2018, July). Introducing big data analytics in high school and college. In <i>Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education</i> (pp. 373374). ACM.: https://doi.org/10.1145/3197091.3205834
R30	Petrova-Antonova, D., Georgieva, O., & Ilieva, S. (2017, June). Modelling of Educational Data Following Big Data Value Chain. In <i>Proceedings of the 18th International Conference on Computer Systems and Technologies</i> (pp. 88–95). ACM.: https://doi.org/10.1145/3134302.3134335
R31	Oi, M., Yamada, M., Okubo, F., Shimada, A., & Ogata, H. (2017). <i>Reproducibility of findings from educational big data</i> . Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference (pp. 536–537). ACM: https://doi.org/10.1145/3027385.3029445
R32	Zhang, M. (2015). Internet use that reproduces educational inequalities: from big data. <i>Computers & Education</i> , 86 (1), 212–223. doi: https://doi.org/10.1016/j.compedu.2015.08.007
R33	Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Muñoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. <i>Computers in Human Behavior</i> , 80 (1), 179–196.: https://doi.org/10.1016/j.chb.2017.11.011
R34	Shorfuzzaman, M., Hossain, M. S., Nazir, A., Muhammad, G., & Alamri, A. (2019). Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. <i>Computers in Human Behavior</i> , 92 (1), 578–588.: https://doi.org/10.1016/j.chb.2018.07.002
R35	Pardos, Z. A. (2017). Big data in education and the models that love them. <i>Current Opinion in Behavioral Sciences</i> , 18 (2), 107–113.: https://doi.org/10.1016/j.cobeha.2017.11.006
R36	Wassan, J. T. (2015). Discovering Big Data Modelling for Educational World. <i>Procedia - Social and Behavioral Sciences</i> , 176, 642–649.: https://doi.org/10.1016/j.sbspro.2015.01.522
R37	Dessi, D., Fenu, G., Marras, M., & Reforgiato Recupero, D. (2019). Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections. <i>Computers in Human Behavior</i> , 92 (1), 468–477.: https://doi.org/10.1016/j.chb.2018.03.004

Table 10 Primary studies references (Continued)

R- ID	References
R38	Selwyn, N. (2014). Data entry: towards the critical study of digital data and education. <i>Learning, Media and Technology</i> , 40 (1), 64–82. doi: https://doi.org/10.1080/17439884.2014.921628
R39	Troisi, O., Grimaldi, M., Loia, F., & Maione, G. (2018). Big data and sentiment analysis to highlight decision behaviours: a case study for student population. <i>Behaviour & Information Technology</i> , 37 (11), 1111–1128. doi: https://doi.org/10.1080/0144929x.2018.1502355
R40	Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016). <i>Big Data Application in Education: Dropout Prediction in Edx MOOCs</i> . Paper presented at the 2016 IEEE Second International Conference on Multimedia Big Data (BigMM) (pp. 440–443). IEEE. doi: https://doi.org/10.1109/BigMM.2016.70

Appendix B

Table 11 Quality assessment criterion

R ID	QA1	QA2	QA3	QA4	Total Score
R1	2	0	2	2	6
R2	2	0	1	2	5
R3	2	2	2	2	8
R4	2	0	1	2	5
R5	2	0	2	2	6
R6	2	2	1	2	7
R7	2	0	2	2	6
R8	2	0	1	2	5
R9	2	0	1	2	5
R10	2	0	2	2	6
R11	2	0	2	2	6
R12	2	0	2	2	6
R13	2	0	2	2	6
R14	2	0	2	2	6
R15	2	0	2	2	6
R16	2	0	2	2	6
R17	2	0	2	2	6
R18	2	0	2	2	6
R19	2	0	2	2	6
R20	2	0	2	2	6
R21	2	2	2	2	8
R22	2	0	2	2	6
R23	2	0	2	2	6
R24	2	0	2	2	6
R25	2	0	2	2	6
R26	2	0	2	2	6
R27	2	0	1	2	5
R28	2	2	2	2	8
R29	2	0	1	2	5
R30	2	2	1	0	5
R31	2	0	1	2	5
R32	2	0	2	1	5
R33	2	2	2	2	8

Table 11 Quality assessment criterion (Continued)

R ID	QA1	QA2	QA3	QA4	Total Score
R34	2	2	2	2	8
R35	2	1	2	0	5
R36	2	2	1	0	5
R37	2	2	1	2	7
R38	2	0	2	1	5
R39	2	2	2	2	8
R40	2	1	0	2	5

Acknowledgements

Not applicable.

Authors' contributions

Maria Ijaz Baig composed the manuscript under the guidance of Elaheh Yadegaridehkordi. Liyana Shuib supervised the project. All authors discussed the results and contributed to the final manuscript.

Funding

Not applicable

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 9 March 2020 Accepted: 10 June 2020

Published online: 02 November 2020

References

- Ahmed, E., Yaqoob, I., Hashem, I. A. T., Shuja, J., Imran, M., Guizani, N., & Bakhsh, S. T. (2018). Recent advances and challenges in mobile big data. *IEEE Communications Magazine*, 56(2), 102–108. China: East China Normal University. <https://doi.org/10.1109/MCOM.2018.1700294>.
- Anshari, M., Alas, Y., & Yunus, N. (2019). A survey study of smartphones behavior in Brunei: A proposal of Modelling big data strategies. In *Multigenerational Online Behavior and Media Use: Concepts, Methodologies, Tools, and Applications*, (pp. 201–214). IGI global.
- Black, P., & William, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>.
- Buffum, P. S., Martinez-Arocho, A. G., Frankosky, M. H., Rodriguez, F. J., Wiebe, E. N., & Boyer, K. E. (2014, March). CS principles goes to middle school: Learning how to teach big data. In *Proceedings of the 45th ACM technical Computer science education*, (pp. 151–156). New York: ACM. <https://doi.org/10.1145/2538862.2538949>.
- Camargo Fiorini, P., Seles, B. M. R. P., Jabbour, C. J. C., Mariano, E. B., & Sousa Jabbour, A. B. L. (2018). Management theory and big data literature: From a review to a research agenda. *International Journal of Information Management*, 43, 112–129. <https://doi.org/10.1016/j.ijinfomgt.2018.07.005>.
- Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J. A., & Muñoz, A. (2019). Analysis of student behavior in learning management systems through a big data framework. *Future Generation Computer Systems*, 90(2), 262–272. <https://doi.org/10.1016/j.future.2018.08.003>.
- Chae, B. K. (2019). A general framework for studying the evolution of the digital innovation ecosystem: The case of big data. *International Journal of Information Management*, 45, 83–94. <https://doi.org/10.1016/j.ijinfomgt.2018.10.023>.
- Chaurasia, S. S., & Frieda Rosin, A. (2017). From big data to big impact: Analytics for teaching and learning in higher education. *Industrial and Commercial Training*, 49(7), 321–328. <https://doi.org/10.1108/ict-10-2016-0069>.
- Chaurasia, S. S., Kodwani, D., Lachhwani, H., & Ketkar, M. A. (2018). Big data academic and learning analytics. *International Journal of Educational Management*, 32(6), 1099–1117. <https://doi.org/10.1108/ijem-08-2017-0199>.
- Coccoli, M., Maresca, P., & Stanganelli, L. (2017). The role of big data and cognitive computing in the learning process. *Journal of Visual Languages & Computing*, 38, 97–103. <https://doi.org/10.1016/j.jvlc.2016.03.002>.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, 65(3), 122–135. <https://doi.org/10.1108/LR-06-2015-0061>.
- Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, (pp. 48–55). San Diego: IEEE. <https://doi.org/10.1109/CTS.2013.6567203>.
- Dessi, D., Fenu, G., Marras, M., & Reforgiato Recupero, D. (2019). Bridging learning analytics and cognitive computing for big data classification in micro-learning video collections. *Computers in Human Behavior*, 92(1), 468–477. <https://doi.org/10.1016/j.chb.2018.03.004>.

- Dinter, B., Jaekel, T., Kollwitz, C., & Wache, H. (2017). *Teaching Big Data Management – An Active Learning Approach for Higher Education*. North America: Paper presented at the proceedings of the pre-ICIS 2017 SIGDSA, (pp. 1–17). North America: AISel.
- Dresner Advisory Services. (2017). Big data adoption: State of the market. ZoomData. Retrieved from <https://www.zoomdata.com/master-class/state-market/big-data-adoption>
- Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in big data and business analytics. *Industrial and Commercial Training*, 47(4), 174–181. <https://doi.org/10.1108/ict-08-2014-0059>.
- Elia, G., Solazzo, G., Lorenzo, G., & Passiante, G. (2018). Assessing learners' satisfaction in collaborative online courses through a big data approach. *Computers in Human Behavior*, 92, 589–599. <https://doi.org/10.1016/j.chb.2018.04.033>.
- Gupta, D., & Rani, R. (2018). A study of big data evolution and research challenges. *Journal of Information Science*, 45(3), 322–340. <https://doi.org/10.1177/0165551518789880>.
- Herschel, R., & Miori, V. M. (2017). Ethics & big data. *Technology in Society*, 49, 31–36. <https://doi.org/10.1016/j.techsoc.2017.03.003>.
- Hirashima, T., Supianto, A. A., & Hayashi, Y. (2017, September). Model-based approach for educational big data analysis of learners thinking with process data. In *2017 International Workshop on Big Data and Information Security (IWBIS)* (pp. 11–16). San Diego: IEEE. <https://doi.org/10.1177/0165551518789880>
- Holland, A. A. (2019). Effective principles of informal online learning design: A theory-building metasynthesis of qualitative research. *Computers & Education*, 128, 214–226. <https://doi.org/10.1016/j.compedu.2018.09.026>.
- Kalaian, S. A., Kasim, R. M., & Kasim, N. R. (2019). Descriptive and predictive analytical methods for big data. In *Web Services: Concepts, Methodologies, Tools, and Applications*, (pp. 314–331). USA: IGI global. <https://doi.org/10.4018/978-1-5225-7501-6.ch018>.
- Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143, 23–37. <https://doi.org/10.1016/j.compag.2017.09.037>.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004), 1–26.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering version 2.3. *Engineering*, 45(4), 13–65.
- Lia, Y., & Zhaia, X. (2018). Review and prospect of modern education using big data. *Procedia Computer Science*, 129(3), 341–347. <https://doi.org/10.1016/j.procs.2018.03.085>.
- Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016). Big Data Application in Education: Dropout Prediction in Edx MOOCs. In *Paper presented at the 2016 IEEE second international conference on multimedia big data (BigMM)*, (pp. 440–443). USA: IEEE. <https://doi.org/10.1109/BigMM.2016.70>.
- Logica, B., & Magdalena, R. (2015). Using big data in the academic environment. *Procedia Economics and Finance*, 33(2), 277–286. [https://doi.org/10.1016/s2212-5671\(15\)01712-8](https://doi.org/10.1016/s2212-5671(15)01712-8).
- Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Muñoz-Gama, J. (2018). Mining theory-based patterns from big data: Identifying self-regulated learning strategies in massive open online courses. *Computers in Human Behavior*, 80(1), 179196. <https://doi.org/10.1016/j.chb.2017.11.011>.
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M. J. (2018). Big Data in Education. In *Paper presented at the proceedings of the sixth international conference on technological ecosystems for enhancing Multiculturalism - TEEM'18, Salamanca, Spain*, (pp. 145–150). New York: ACM. <https://doi.org/10.1145/3284179.3284206>.
- Mikaléf, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and e-Business Management*, 16(3), 547–578. <https://doi.org/10.1007/10257-017-0362-y>.
- Mohammadpoor, M., & Torabi, F. (2018). Big Data analytics in oil and gas industry: An emerging trend. *Petroleum*. In press. <https://doi.org/10.1016/j.petlm.2018.11.001>.
- Muthukrishnan, S. M., & Yasin, N. B. M. (2018). Big Data Framework for Students' Academic. Paper presented at the symposium on computer applications & industrial electronics (ISCAIE), Penang, Malaysia (pp. 376–382). USA: IEEE. <https://doi.org/10.1109/ISCAIE.2018.8405502>
- Neilson, A., Daniel, B., & Tjandra, S. (2019). Systematic review of the literature on big data in the transportation Domain: Concepts and Applications. *Big Data Research*. In press. <https://doi.org/10.1016/j.bdr.2019.03.001>.
- Nelson, M., & Pouchard, L. (2017). A pilot "big data" education modular curriculum for engineering graduate education: Development and implementation. In *Paper presented at the Frontiers in education conference (FIE), Indianapolis, USA*, (pp. 1–5). USA: IEEE. <https://doi.org/10.1109/FIE.2017.8190688>.
- Nie, M., Yang, L., Sun, J., Su, H., Xia, H., Lian, D., & Yan, K. (2018). Advanced forecasting of career choices for college students based on campus big data. *Frontiers of Computer Science*, 12(3), 494–503. <https://doi.org/10.1007/s11704-017-6498-6>.
- Oi, M., Yamada, M., Okubo, F., Shimada, A., & Ogata, H. (2017). Reproducibility of findings from educational big data. In *Paper presented at the proceedings of the Seventh International Learning Analytics & Knowledge Conference*, (pp. 536–537). New York: ACM. <https://doi.org/10.1145/3027385.3029445>.
- Ong, V. K. (2015). Big Data and Its Research Implications for Higher Education: Cases from UK Higher Education Institutions. In *Paper presented at the 2015 IIAI 4th international confress on advanced applied informatics*, (pp. 487–491). USA: IEEE. <https://doi.org/10.1109/IIAI-AAI.2015.178>.
- Ozgur, C., Kleckner, M., & Li, Y. (2015). Selection of statistical software for solving big data problems. *SAGE Open*, 5(2), 59–94. <https://doi.org/10.1177/2158244015584379>.
- Pardos, Z. A. (2017). Big data in education and the models that love them. *Current Opinion in Behavioral Sciences*, 18(2), 107–113. <https://doi.org/10.1016/j.cobeha.2017.11.006>.
- Petrova-Antonova, D., Georgieva, O., & Ilieva, S. (2017, June). Modelling of educational data following big data value chain. In *Proceedings of the 18th International Conference on Computer Systems and Technologies* (pp. 88–95). New York City: ACM. <https://doi.org/10.1145/3134302.3134335>
- Qiu, R. G., Huang, Z., & Patel, I. C. (2015, June). A big data approach to assessing the US higher education service. In *2015 12th International Conference on Service Systems and Service Management (ICSSSM)* (pp. 1–6). New York: IEEE. <https://doi.org/10.1109/ICSSSM.2015.7170149>
- Ramos, T. G., Machado, J. C. F., & Cordeiro, B. P. V. (2015). Primary education evaluation in Brazil using big data and cluster analysis. *Procedia Computer Science*, 55(1), 10311039. <https://doi.org/10.1016/j.procs.2015.07.061>.

- Rimmon-Kenan, S. (1995). What Is Theme and How Do We Get at It?. *Thematics: New Approaches*, 9–20.
- Roy, S., & Singh, S. N. (2017). Emerging trends in applications of big data in educational data mining and learning analytics. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, (pp. 193–198). New York: IEEE. <https://doi.org/10.1109/confluence.2017.7943148>.
- Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, 54(5), 758–790. <https://doi.org/10.1016/j.ipm.2018.01.010>.
- Santoso, L. W., & Yulia (2017). Data warehouse with big data Technology for Higher Education. *Procedia Computer Science*, 124(1), 93–99. <https://doi.org/10.1016/j.procs.2017.12.134>.
- Sedkaoui, S., & Khelifaoui, M. (2019). Understand, develop and enhance the learning process with big data. *Information Discovery and Delivery*, 47(1), 2–16. <https://doi.org/10.1108/idd-09-2018-0043>.
- Selwyn, N. (2014). Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology*, 40(1), 64–82. <https://doi.org/10.1080/17439884.2014.921628>.
- Shahat, O. A. (2019). A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, 91(1), 620–633. <https://doi.org/10.1016/j.future.2018.06.046>.
- Shorfuzzaman, M., Hossain, M. S., Nazir, A., Muhammad, G., & Alamri, A. (2019). Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. *Computers in Human Behavior*, 92(1), 578–588. <https://doi.org/10.1016/j.chb.2018.07.002>.
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>.
- Sledgianowski, D., Goma, M., & Tan, C. (2017). Toward integration of big data, technology and information systems competencies into the accounting curriculum. *Journal of Accounting Education*, 38(1), 81–93. <https://doi.org/10.1016/j.jaccedu.2016.12.008>.
- Sooriamurthi, R. (2018). Introducing big data analytics in high school and college. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (pp. 373–374). New York: ACM. <https://doi.org/10.1145/3197091.3205834>.
- Sorensen, L. C. (2018). "Big data" in educational administration: An application for predicting school dropout risk. *Educational Administration Quarterly*, 45(1), 1–93. <https://doi.org/10.1177/0013161x18799439>.
- Su, Y. S., Ding, T. J., Lue, J. H., Lai, C. F., & Su, C. N. (2017). Applying big data analysis technique to students' learning behavior and learning resource recommendation in a MOOCs course. In *2017 International conference on applied system innovation (ICASI)* (pp. 1229–1230). New York: IEEE. <https://doi.org/10.1109/ICASI.2017.7988114>.
- Troisi, O., Grimaldi, M., Loia, F., & Maione, G. (2018). Big data and sentiment analysis to highlight decision behaviours: A case study for student population. *Behaviour & Information Technology*, 37(11), 1111–1128. <https://doi.org/10.1080/0144929x.2018.1502355>.
- Ur Rehman, M. H., Yaqoob, I., Salah, K., Imran, M., Jayaraman, P. P., & Perera, C. (2019). The role of big data analytics in industrial internet of things. *Future Generation Computer Systems*, 92, 578–588. <https://doi.org/10.1016/j.future.2019.04.020>.
- Veletsianos, G., Reich, J., & Pasquini, L. A. (2016). The Life Between Big Data Log Events. *AERA Open*, 2(3), 1–45. <https://doi.org/10.1177/2332858416657002>.
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13. <https://doi.org/10.1016/j.techfore.2015.12.019>.
- Wassan, J. T. (2015). Discovering big data modelling for educational world. *Procedia - Social and Behavioral Sciences*, 176, 642–649. <https://doi.org/10.1016/j.sbspro.2015.01.522>.
- Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming—a review. *Agricultural Systems*, 153, 69–80. <https://doi.org/10.1016/j.jagsy.2017.01.023>.
- Wu, P. J., & Lin, K. C. (2018). Unstructured big data analytics for retrieving e-commerce logistics knowledge. *Telematics and Informatics*, 35(1), 237–244. <https://doi.org/10.1016/j.tele.2017.11.004>.
- Xu, L. D., & Duan, L. (2019). Big data for cyber physical systems in industry 4.0: A survey. *Enterprise Information Systems*, 13(2), 148–169. <https://doi.org/10.1080/17517575.2018.1442934>.
- Yang, F., & Du, Y. R. (2016). Storytelling in the age of big data. *Asia Pacific Media Educator*, 26(2), 148–162. <https://doi.org/10.1177/1326365x16673168>.
- Yassine, A., Singh, S., Hossain, M. S., & Muhammad, G. (2019). IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems*, 91(2), 563–573. <https://doi.org/10.1016/j.future.2018.08.040>.
- Zhang, M. (2015). Internet use that reproduces educational inequalities: Evidence from big data. *Computers & Education*, 86(1), 212–223. <https://doi.org/10.1016/j.compedu.2015.08.007>.
- Zheng, M., & Bender, D. (2019). Evaluating outcomes of computer-based classroom testing: Student acceptance and impact on learning and exam performance. *Medical Teacher*, 41(1), 75–82. <https://doi.org/10.1080/0142159x.2018.1441984>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.