Lnenicka et al. International Journal of Educational Technology in Higher Education (2020) 17:28
https://doi.org/10.1186/s41239-020-00208-z

International Journal of Educational Technology in Higher Education

RESEARCH ARTICLE Open Access

Big and open linked data analytics: a study on changing roles and skills in the higher educational process

Martin Lnenicka*, Hana Kopackova, Renata Machova and Jitka Komarkova

* Correspondence: martin.lnenicka@gmail.com
Faculty of Economics and Administration, University of Pardubice, Studentska 95, 53210 Pardubice, Czech Republic

Abstract

The concept of openness and information sharing (linking) together with increasing amounts of data available significantly affect the current educational system. Institutions as well as other stakeholders are facing challenges how to successfully deal with them and potentially profit from them. In this regard, this paper explores opportunities of big and open linked data analytics in the educational process intended to develop the new set of skills. A comprehensive literature review resulted in a framework of relevant skills, namely soft, hard, and data analytics skills. Their importance was evaluated using a Delphi method. In order to determine the relationships between involved stakeholders, their roles and requirements, a stakeholder theory is utilized. It resulted in the identification of current and emerging roles of stakeholders in the data analytics ecosystem. A structural classification of stakeholders' influences and impacts then represents a necessary background for establishing strategies for the development of the right skills needed to gain the value from these data. This paper provides a comprehensive view on big and open linked data analytics in the educational context, defines and interlinks data-related with current roles as well as the skills required to perform data analytics.

Keywords: Analytics, Higher education, Skills, Delphi method, Stakeholder theory, Framework

Introduction

The whole education sector is facing increasing amounts of data available, but lack to utilize such opportunities for working with them (Cervone, 2016; Huda et al., 2017; Máchová, Komárková, & Lněnička, 2016; Picciano, 2012; Vieira, Parsons, & Byrd, 2018). New tools and techniques for storing, managing, processing, analysing, visualization, and sharing of these data fast and easily are available for everyone (Aghabozorgi, Mahroeian, Dutt, Wah, & Herawan, 2014; Khriyenko & Khriyenko, 2013; Klašnja-Milićević, Ivanović, & Budimac, 2017). This process is called data analytics and refers to the analysis of information from a particular domain (Zadeh, Schiller, Duffy, & Williams, 2018). However, due to a variety of systems and environments, data generated in these contexts are

difficult to understand and interpret with traditional data analytics skills (Aikat et al., 2017; Chatti, Muslim, & Schroeder, 2017; Demchenko, Gruengard, & Klous, 2014).

New technological layers, frameworks, and models that are capable of capturing, storing, managing, visualizing and processing these data are needed to support the educational process (Chatti et al., 2017; Huda et al., 2017; Liñán & Pérez, 2015; Macfadyen, Dawson, Pardo, & Gasevic, 2014; Máchová et al., 2016; Vieira et al., 2018). At the same time, many of learning materials as well as courses and lectures are open and available anywhere as Open Educational Resources (OER), mostly in the form of interactive (tutorial) courseware and Massive Open Online Courses (MOOC), which combine different OER, e-learning methods and social networks culminating in an online learning experience. Such evolution influences today's educational systems and requires to build data-centred skills (Aikat et al., 2017; Colpaert, 2018; Demchenko et al., 2017; Khriyenko & Khriyenko, 2013; Klašnja-Milićević et al., 2017; Li & Ni, 2015; Mikroyannidis, Domingue, Maleshkova, Norton, & Simperl, 2016). It changes learning needs and requires a reorientation toward the development of novel approaches and advancements in higher educational institutions (Gupta, Goul, & Dinter, 2015; Sedkaoui, 2018). These demands then create the need for systemic change in education focusing on preparing new specialists (data-related roles) and training for new skills (Coccoli, Maresca, & Stanganelli, 2017; Demchenko et al., 2014; Demchenko et al., 2017). According to the findings of the recent study of Mikalef, Giannakos, Pappas, and Krogstie (2018), there is a gap between the skills needed and the ones taught in academic curricula.

Due to these many new uses and interactions, a new concept called Big and Open Linked Educational Data (BOLED) can be introduced. It origins from the intersection of big and open and linked data analytics in the public sector (Janssen & Kuk, 2016; Lněnička & Komárková, 2019). However, these challenges are also faced by other sectors such as educational institutions. In this regard, BOLED analytics can is defined as "*the acquisition and extraction, management and preparation, storing and archiving, processing and analysis, visualization and use, and publication, sharing and reuse of data about contexts of learners, educators and other involved stakeholders, for purposes of understanding and improving teaching and learning processes as well as the environments in which they occur*" (Lněnička, Máchová, Komárková, & Čermáková, 2018). In our study, we use this concept as an umbrella term that enables us to involve all relevant stakeholders and activities into a single ecosystem in which all can benefit from working with these data.

Therefore, the aim of this paper is to review good practices in the use of BOLED analytics as a basis for a stakeholder-based research that can contribute to the development of stakeholders' skills and field-specific competencies in academic curricula. This perspective is crucial to understand how stakeholders can benefit from reusing these data and create both economic and social value out of the available data sources. In particular, we provide a set of skills that stakeholders need to focus in order to reuse BOLED and a description of their roles. This enables us to introduce a structured framework serving as an input into the evaluation process using the Delphi method. The results contribute to practice by providing a comprehensive view on what skills are required to perform BOLED analytics and work with these data. Knowing what is required for each role and data-based activity can help institutions in focusing on the right skillsets as well as develop appropriate courses. Furthermore, a newly proposed framework focused

on the skills of involved stakeholders is expected to overcome these barriers and risks associated with BOLED analytics.

This paper exploits several questions about the potential of BOLED analytics in the educational process intended to develop the new set of skills. Our research is focused on these questions:

RQ1: How BOLED analytics can help and stimulate the educational process?
RQ2: Who are the stakeholders involved in the process and what are relationships between their roles?
RQ3: What skills are most important for stakeholders in different roles in order to reuse and create value from BOLED?
RQ4: What could be done to address these challenges and improve this process in the future?

We adopt a research methodology consisting of a two-step strategy to investigate these questions. First, a comprehensive literature review is conducted to identify: 1) the background and benefits of BOLED in the educational process; 2) the stakeholders involved in the process and 3) key skills that are shaped by requirements of BOLED. The second step involves performing the Delphi process by surveying experts to measure the diversity of their opinions on this topic that spans a wide range of disciplines (Okoli & Pawlowski, 2004). In order to understand and formalize the relationships of involved stakeholders together with their influences and impacts, we address these issues through an alignment of concepts from the stakeholder theory (Rowley, 1997) in higher education and its communities (Jongbloed, Enders, & Salerno, 2008). The results are quantified for required skills as well as current and emerging roles of stakeholders.

This paper is structured as follows. This introduction is followed by Section 2 describing the research methodology. Subsequently, literature review and the key concepts of this paper are presented in Section 3. The following Section 4 provides the background on developing the framework. Section 5 is devoted to results and most important findings. These are discussed against the results made by other authors in Section 6. In Section 7 our concern is to present recommendations, consider limitations of our study, and look towards future work. Section 8 concludes the paper.

Research methodology

The research gaps are summarized in Fig. 1. It reflects the need to deal with increasing data volumes in the educational process on the one hand and enable involved stakeholders to efficiently create value from these data using relevant hard, soft, and BOLED analytics skills on the other. The gaps are formalized in the higher education context of the faculty focused on economics, public administration, and computer science. There are identified the following gaps: a) a necessity to identify a set of required skills for BOLED analytics; b) a necessity of an adequate framework. The paper addresses the firs gap. It aims at identification of a set of skills, which is necessary for stakeholders in higher education to improve the educational process through analyses of available big and open and linked data.

Since research on stakeholders' skills to reuse data analytics and create value from these data is still lacking, a new framework is necessary to identify relevant skills. In

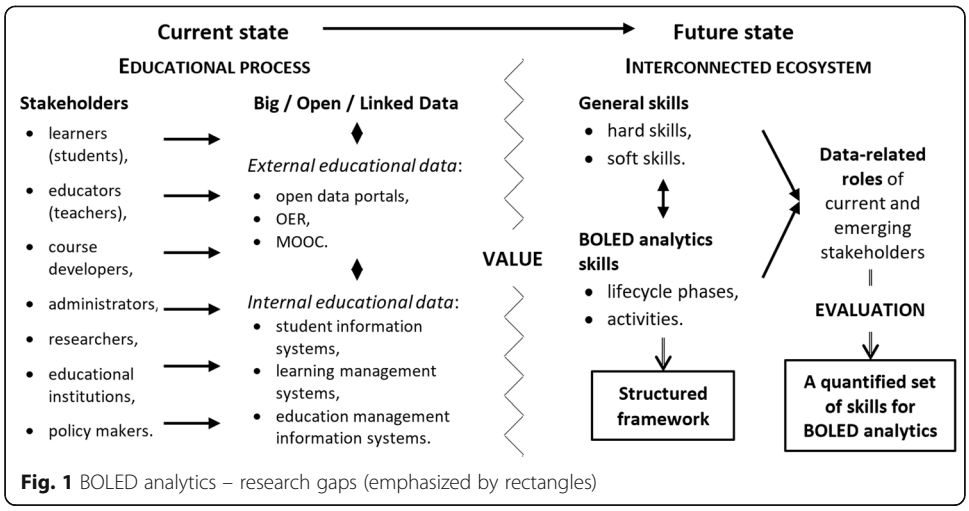

Fig. 1 BOLED analytics – research gaps (emphasized by rectangles)

order to help various groups of stakeholders to deal with BOLED analytics, the following methodology is introduced. First, a comprehensive literature review is done. It contains a cross-search among several databases to retrieve related papers. Among others, it consists of reviewing the literature related to the identification of involved stakeholders, definition of the BOLED analytics lifecycle, and relevant skills. The process of the comprehensive literature review consists of 1) search for key words (limited to educational research); 2) elimination of duplicate papers; 3) selection of papers based on their titles and abstracts; and 4) addition of further papers based on recommendations of experts. Aiming to understand the multiple requirements ascribed to BOLED in the educational context, we need to analyse who is involved with BOLED in that context and what their particular perspectives are. For this purpose, this paper utilizes the stakeholder theory.

As can be seen in Fig. 2, the application of the Delphi method starts with developing the framework and it is further used to evaluate the framework. The main reason for the use of this method is the lack of other research evidence on the topic. Further arguments include: 1) it does not need to bring the experts together and enables to involve more experts with a broader interdisciplinary background, 2) it affords anonymity to participants and privacy for iteration and to change one's mind over several rounds and 3) it facilitates the transformation of opinions into consensus. The steps of this method follow the guidelines of Okoli and Pawlowski (2004). The required qualifications of experts include a minimum of five years' experience in the field of the research problem and at least B.Sc. A group of eight experts have been specially selected for their particular expertise on the topic. It offers the opportunity to check the validity of the cross-disciplinary nature of the issue (Grisham, 2009).

Experts' qualifications are summarized in Table 1, including a job position, academic title, years of experience, and professional expertise. Our aim was to involve experts that represent a range of opinions about the research topic and are able to evaluate the relevant skills. All the experts are familiar with the Delphi process and agreed to participate in all Delphi rounds needed. Their motivation as well as the quality of the instructions are important since a high dropout rate among experts may lead to non-representative results (Grisham, 2009; Okoli & Pawlowski, 2004). Ensuring that the

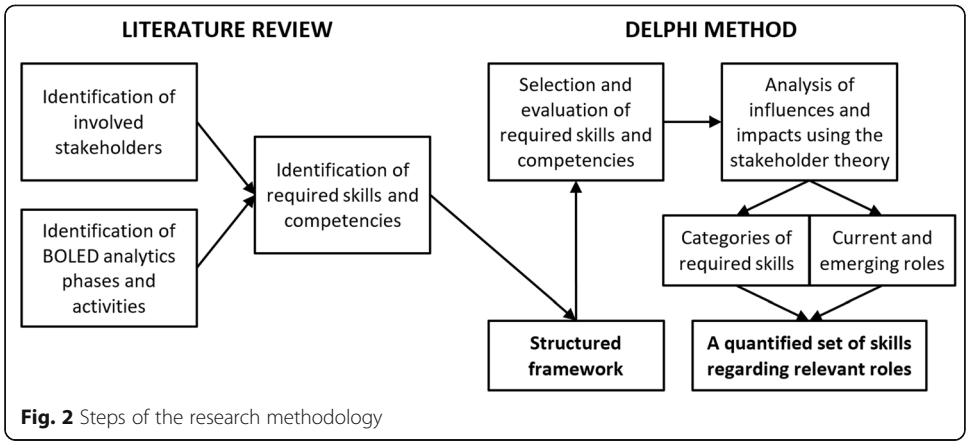

Fig. 2 Steps of the research methodology

process is robust, consistent, transparent, and meets the quality aspects arising from the previous uses of the Delphi method determines the external validity of the results.

The literature review is done to provide the basis for categorizing the key skills, which are then presented to the panel of experts. In the qualitative first round, the experts are asked to identify the range of required skills and classify them into categories. This list is used to construct the online survey instrument distributed in subsequent quantitative rounds. After each of these rounds, the responses are provided to all panel members and they are asked to decide whether they want to change any of their responses. The subsequent round questionnaire is constructed from the results gathered from the previous one. The process is finished after no one wants to change their responses. In this study, the Delphi process is limited to a maximum of five rounds. However, only three rounds were realized. A five-point Likert scale, from 1 = extremely unimportant to 5 = extremely important, is utilized to determine the suitability of selected skills for each role in which the stakeholder may work with BOLED. The scores of the panel members are analysed after each round to calculate mean values and standard deviations of each question. Mean value indicates the central tendency of feedback, whereas standard deviation shows the achievement degree of consensus.

Literature review

Three electronic databases were searched to collect an input set of papers: Web of Science, Scopus, and IEEE. The search covered the period of 2008–2020. Reviewed journals,

Table 1 Experts' qualifications participated in the Delphi process

Job position	Academic title	Experience	Expertise
Academic staff	Assoc. Prof.	> 20	Computer science and research
Academic staff	Ph.D.	15–20	Economics, management and research
Academic staff	Ph.D.	10–15	Administration and research
Secondary school teacher	Ph.D.	5–10	Computer science and pedagogy
Secondary school teacher	M.Sc.	> 20	Pedagogy and mathematics
Data analyst	M.Sc.	5–10	Computer science and engineering
Head of the department of education	M.Sc.	> 20	Management and administrative law
Human resources manager	B.Sc.	15–20	Social science and human resource management

conference proceedings, and book chapters were included. Used combinations of key words and number of results are described by Table 2. The search resulted into 378 results.

Afterwards, duplicate papers were removed, reducing the sample to 182 papers with full text availability, see Fig. 3. The next search phase was focused on academic data analytics including recommendations and guidelines on how to build relevant skills and prepare stakeholders for the challenges of BOLED analytics. Further papers were added based on recommendations of experts. This search led to 72 unique documents, which were analysed by the authors.

Open data for education and interlinking educational data

Educational activities are nowadays happening in open and networked learning environments, characterized by increasing complexity and fast-paced change (Chatti et al., 2017).

Table 2 Results of the search through databases

Key words combinations	Web of Science		Scopus			IEEE	
	proceedings	journals	proceedings	journals	book chapters	proceedings	journals
big educational data analytics AND stakeholders	6	3	8	6	1	9	0
big educational data analytics AND roles	11	18	15	12	2	10	1
big educational data analytics AND skills	15	11	18	11	2	14	0
big educational data analytics AND ecosystem	4	3	6	4	1	6	0
open educational data analytics AND stakeholders	4	4	12	3	0	6	2
open educational data analytics AND roles	6	12	13	9	1	3	0
open educational data analytics AND skills	12	9	14	11	3	7	2
open educational data analytics AND ecosystem	3	2	3	3	1	6	0
linked educational data analytics AND stakeholders	0	1	0	1	0	1	1
linked educational data analytics AND roles	1	3	0	2	0	0	0
linked educational data analytics AND skills	2	2	1	1	0	3	0
linked educational data analytics AND ecosystem	0	3	1	0	0	3	0
big and open linked educational data analytics AND stakeholders	0	0	0	0	0	0	0
big and open linked educational data analytics AND roles	0	1	0	0	0	0	0
big and open linked educational data analytics AND skills	0	1	1	0	0	0	0
big and open linked educational data analytics AND ecosystem	0	0	0	0	0	1	0
Sum	**64**	**73**	**92**	**63**	**11**	**69**	**6**

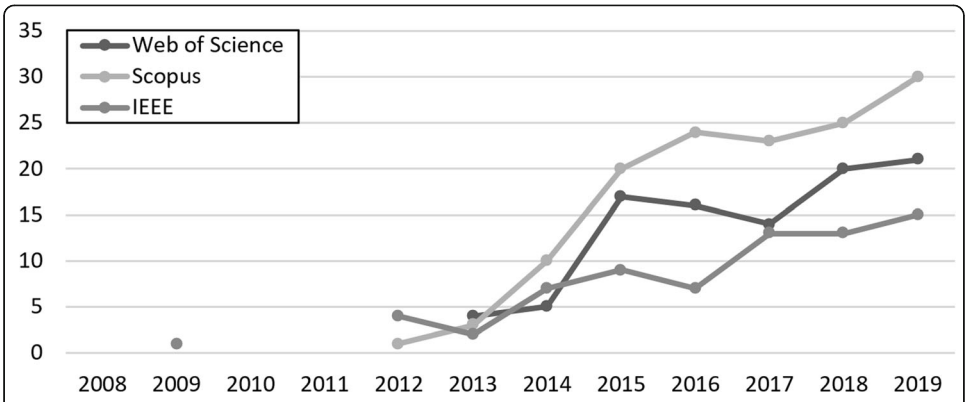

Fig. 3 Number of papers in particular databases by year

The increasing availability of OER, MOOC and other open data sources provides challenges for stakeholders to get more from these data (Atenas, Havemann, & Priego, 2015; Colpaert, 2018; Khriyenko & Khriyenko, 2013; Zuiderwijk, Janssen, & Davis, 2014). Verbert, Manouselis, Drachsler, and Duval (2012) reported that the availability of open data sets is considered as key for research and application purposes. Open Education (OE) is a philosophy and a movement about the way people should produce, share, and build on knowledge (Colpaert, 2018). While OER are freely accessible, openly licensed documents and media that are useful for teaching, learning, research, and assessing (Navarrete & Luján-Mora, 2015), open data provide a large sources of information on a wide range of subjects (Van der Waal et al., 2014). Janssen, Charalabidis, and Zuiderwijk (2012) defined them as "*non-privacy-restricted and non-confidential data which are produced with public money and are made available without any restrictions on its usage or distribution.*" Since open data are meant to be freely used and reused, not all educational data can be published as open data, mostly due to security and privacy restrictions.

The OE refers to the advancement of education through "*open technology, open content and open knowledge*" (Iiyoshi & Kumar, 2008) and its aim is to serve the greater public good through the sharing of topical and thematic learning objects as well as intact course materials and curricula. This OE relies on open data sources. However, open data are not always OER. They become OER when used within pedagogical contexts. On the contrary, OER are not always open data, they become open data when they are published with open license and comply with other principles of open data (Atenas, & Havemann, L, 2015; Atenas et al., 2015; Colpaert, 2018). According to Khriyenko and Khriyenko (2013), the next generation of innovative education environments will apply the achievements of the open data initiative and move towards learner-driven society-oriented systems. Chatti et al. (2017) presented an open learning analytics ecosystem that aims at supporting learning and teaching in fragmented, diverse, and networked learning environments.

However, there are still large amounts of data that need to be centralized in order to create a central point of access to BOLED. Since data are useful only when they are being used and transformed into knowledge, publishing data on data portals enables various stakeholders to add their data, open them for everyone and provide analytical opportunities (Eldridge, Hobbs, & Moran, 2018; Kyritsi, Zorkadis, Stavropoulos, & Verykios, 2019; Lněnička & Máchová, 2015). According to Khriyenko and Khriyenko (2013), open data

portals could enable the new generation of learners and education-related content providers to create, share, search, combine and deliver trustworthy and competent information easily.

The context of reusing these data in education is being brought to the attention of researchers and practitioners. It is characterized by the active engagement of stakeholders through cooperation, collaboration, and participation activities on data resulting in an improvement of the educational process. In this regard, new important insights on the interactions of the stakeholders can be gained by performing data analytics on these data (Eckartz, van den Broek, & Ooms, 2016; Huda et al., 2017). The emergence of OER has facilitated online education through the use and sharing of a wide variety of quality learning materials available through open services provided on the web (Mikroyannidis et al., 2016). These data sets can be used by educators as OER to support different teaching and learning activities, allowing learners to gain experience working with the same raw data researchers and policy-makers generate and use (Khriyenko & Khriyenko, 2013). In this way, educators can facilitate learners to understand how information is generated, processed, analysed and interpreted (Atenas et al., 2015). The reuse of open data in education can reduce the time and money of stakeholders, make educational platforms and tools more efficient, flexible, and accessible for new groups of users. It especially provides stakeholders more options to manage the learning process, find solutions to complex problems and create new products and services from these data (Atenas, & Havemann, L, 2015; Iiyoshi & Kumar, 2008; Janssen et al., 2012).

Linked data refer to a set of best practices for publishing and connecting structured data allowing data providers to publish and link their data with other information sources over the Web into a global information space containing billions of assertions, i.e. the Web of Data (Heath & Bizer, 2011; Van der Waal et al., 2014). Linked data are mostly taken together with open data as Linked Open Data (LOD) because interlinking data provides more context and greater opportunities for stakeholders to reuse and exploit them (Colpaert, 2018; Janssen & Kuk, 2016). Geiger and von Lucke (2012) defined LOD as: "*all stored data connected via the World Wide Web which could be made accessible in a public interest without any restrictions for usage and distribution*." The more stakeholders can access relevant data sets as LOD, e.g. through new functionalities, such as APIs or other interface, provided by learning management systems, the more value and enrichment they create (Colpaert, 2018).

While there are big amounts of OER available today, finding, querying, and integrating, interlinking these resources is often difficult, Navarrete and Luján-Mora (2015) proposed the use of a 5-star linked open data scheme to enhance the use of OER. They claimed that the adoption of linked data technology to connect and enrich OER environments requires the recognition of both perspectives, technical and end-users. A methodology for the design and implementation of an educational curriculum about LOD, supported by multimodal OER was presented by Mikroyannidis et al. (2016). Khriyenko and Khriyenko (2013) reported the need for mashup-based platforms enabling a new content creation based on automated/semi-automated composition of content retrieved from external sources. These tools should enhance content creation process and produce innovative study materials. A framework reflecting various dimensions for the collection and sharing of educational data sets was presented by Verbert

et al. (2012). Publishing these data as LOD helps to overcome the issue of privacy rights and licensing of educational data as well as provides a standard set of guidelines and parameters for data formats. Colpaert (2018) emphasized personalization (adaptation of difficulty level, task type, etc. to the learner) and contextualization (adaption of content to the geotemporal location of the learner) of the learning process. Both these approaches depend on BOLED and fit within current phenomena such as augmented reality, smart cities, and the Internet of Things (IoT) (Colpaert, 2018).

Big data in education and analytics over these data

According to authors such as Aguilar (2018), Demchenko et al. (2014), Gkontzis, Kotsiantis, Panagiotakopoulos, and Verykios (2019), Klašnja-Milićević et al. (2017); Liñán and Pérez (2015) and Picciano (2012), technological progress is driving the research on data mining and learning analytics to the era of big data in which different technologies are overlapping and complementing each other to provide insights into the learning process. The spread of these new technologies together with online courses and data portals provides new opportunities, approaches and challenges for data analytics (Jena, 2019; Klašnja-Milićević et al., 2017; Macfadyen et al., 2014; Pecori, 2018; Romero & Ventura, 2010). Similarly, new methods and approaches can be implemented to gain new insights from data and make education content more attractive and motivational for learners, educators and the whole education ecosystem (Cen, Ruta, & Ng, 2015; Gkontzis et al., 2019; Huda et al., 2017; Khriyenko & Khriyenko, 2013).

Miller (2014) reported that every profession, whether business or technical, will be impacted by big data and analytics. The educational process, which nowadays mostly relies on web and computer-based education, is producing vast amounts of data about activities of involved stakeholders (Aghabozorgi et al., 2014; Atenas et al., 2015; Coccoli et al., 2017; Zheng et al., 2014). These big educational data are characterized by the increasing volume, variety, and velocity of data streams and variability, veracity, and value of them (Demchenko et al., 2014; Pecori, 2018; Picciano, 2012). Other characteristics may include: validity (applicable), volatility (temporal), verbosity (text), verification (trust), visualisation (presentation), or vulnerability (security and assurance) (Self, 2014). As reported by Demchenko et al. (2014), in contrary to more focused technology domain such as cloud computing, big data fuse computer technology, data analytics and research methods – the three domains that earlier have been considered as rather independent.

Big data provide *"a technical platform to well meet the need of integrated learning and teaching process in a systematic perspective."* (Li & Ni, 2015) and underpin the development of educational technologies (Aguilar, 2018). According to Fernández, Peralta, Benítez, and Herrera (2014), a cloud platform is a natural structure for the implementation of data mining techniques and their application to growing big educational data. It also eases the implementation of data mining techniques to work in a distributed scenario, regarding the large databases generated from e-learning. For example, Hashem et al. (2015) described the use of cloud computing in big data as follows: "*Large data sources from the cloud and Web are stored in a distributed fault-tolerant database and processed through a programing model for large data sets with a parallel distributed algorithm in a cluster*." Lněnička and Komárková (2015) introduced a list of platforms, tools, and services related to this issue. The other paradigm, called fog computing, reducing the amount of data that need

to be transported to the cloud for processing, analysis and storage of educational big data streams was developed by Pecori (2018). The introduction of big data analytics tools in educational institutions and ensuring that they can be safely incorporated into their technical and operational infrastructures was explored by Sedkaoui (2018) and Self (2014), respectively. Yu and Wu (2015) introduced relevant application examples of educational data mining to show how big data in education help to solve educational issues, focusing on different groups of involved stakeholders and their objectives.

Data in different formats and/or from different sources have to be analysed and interpreted to support data-driven decision making (Miller, 2014; Pecori, 2018; Picciano, 2012) and to construct knowledge (Aikat et al., 2017; Atenas et al., 2015; Huda et al., 2017; Verbert et al., 2012). According to Self (2014), two developing areas of using BOLED analytics in higher educational institutions are the analysis of social media for sentiment analysis and the evaluation and prediction of student (non) achievement and progression. Other uses involve administrative applications such as recruitment and admissions processing, registrar activities, financial planning or donor tracking (Picciano, 2012), the individualization and personalization of the learning experience (Aguilar, 2018; Cen et al., 2015; Chatti et al., 2017; Gkontzis et al., 2019). Aghabozorgi et al. (2014) distinguishes between educational data analytics and learning analytics dealing with predictive models in educational systems. Insights into improving learners' skills through the use of big data analytics tools were discussed by Sedkaoui (2018). Dealing with managerial and organizational concerns, Cervone (2016) provided a list of eight aspects to be considered in the development of a data analytics strategy in organizations.

A model approach to process big educational data from the Moodle system in the cloud using the Apache Hadoop cluster was proposed by Máchová et al. (2016). Their results suggested that the approach has a potential to be widely used for the processing of big educational data, especially with user interaction data. Marjanovic, Milovanovic, and Radenkovic (2014) introduced a big data infrastructure deployed as Hadoop platform in order to improve educational process using data generated through users' online activities. The platform can be also integrated with the learning management system Moodle. The Hadoop ecosystem and the Moodle system were also used by Gkontzis et al. (2019), who aimed to predict student's attrition in three target categories, or Jena (2019), who applied different machine learning techniques to understand students' sentiments. Models encompassing the key phases of educational data lifecycle with respect to understanding of big data analytics within the teaching and learning process were introduced by Chatti et al. (2017), Huda et al. (2017) and Klašnja-Milićević et al. (2017). In order to help users search in big educational data, Abdelouarit, Sbihi, and Aknin (2015) proposed a tool called Big-Learn that integrates the mix of structured and unstructured data into one data layer to facilitate access and more relevant search according to the expectations of the learner. Pecori (2018) demonstrated that big data stream mining and fog computing applied to distance learning environments may foster many benefits as well as improving some drawbacks of already presented cloud-based e-learning infrastructures.

In order to work with and gain from BOLED, it is important to ensure that data skills are able to support them (Atenas, & Havemann, L, 2015), including architectures that are able to integrate the relevant technologies and components (Huda et al., 2017; Pecori, 2018). As reported by Eckartz et al. (2016), it is needed to focus on: (1) infrastructure and enabling technologies that include all hardware and software needed to manage data lifecycle, (2) an IT strategy that includes planning, data management and

governance, and measures to secure data, and (3) interoperability that includes skills on working with standards and ability to integrate information systems. Providing the theoretical and practical knowledge in the educational process, Sedkaoui (2018) emphasized the importance of tools to store large volumes of data in scaled-up architectures, integrate of various types of data, real-time data processing, Web data analysis, and data quality management.

Finally, it should be also noted that there is an increasing interest in smart technologies and their applications in BOLED analytics (Coccoli et al., 2017; Lněnička, Máchová, Komárková, & Pásler, 2017). In such a context, smart technologies are characterized by "*a transition to some extent intelligent elec-tronic devices or systems enabling broad access to relevant information and knowledge that should help people to get informed and improve the decision-making process*" (Lněnička & Komárková, 2019). Coccoli et al. (2017) introduced a model of a smart university in which the exploitation of big data and cognitive computing systems redesign the learning process and enable knowledge to grow rapidly and easy to share among both educators and learners.

Thus, BOLED analytics can add value to teaching and learning practices by providing greater insights into the learning processes, identify the impact of various learning strategies, improve learners' experience, provide more effective evidence-based decision making and collaborative educational environment (Aguilar, 2018; Cen et al., 2015; Coccoli et al., 2017; Fernández et al., 2014; Jena, 2019; Klašnja-Milićević et al., 2017; Macfadyen et al., 2014; Picciano, 2012). In addition, observing the way learners perform in the e-learning systems helps to assess the efficiency of these systems and improve instructional materials (Aghabozorgi et al., 2014; Cen et al., 2015; Fernández et al., 2014; Marjanovic et al., 2014; Zheng et al., 2014). The deployment of big data practices is also intended to accelerate the governing by numbers, i.e. making the collection of educational data, its processes of calculation and its consequences into an automated, real-time and recursive process (Williamson, 2016). On the other hand, new skills are required to gain valuable information and a deeper insight into the processes taking place around educational data analytics (Atenas et al., 2015; Demchenko et al., 2017; Máchová et al., 2016).

Developing the framework

Identifying stakeholders and their roles in the BOLED analytics ecosystem

Identification of stakeholders in the BOLED analytics ecosystem is based on a theory of stakeholder influences constructed by Rowley (1997), which "*accommodates multiple, interdependent stakeholder demands and predicts how organizations respond to the simultaneous influence of multiple stakeholders.*" It is important since the explanation of how an organization functions with respect to the relationships and influences existing in the ecosystem provide boundaries for stakeholders' skills analysis. As reported by Rowley (1997), the multiple and interdependent interactions that simultaneously exist in stakeholder environments have to be considered to predict organizational responses to these forces. In this regard, organizations must answer the simultaneous demands of multiple stakeholders and take into account their influences and impacts on the organization.

In order to meet the requirements of BOLED analytics, it is important to enhance the ecosystem of stakeholders in the educational institutions. Although the learners

and educators are still the key roles in the educational process, the pressure is on their skills. The research on this topic is mostly dealing with existing stakeholders and how they can profit from these new approaches and technologies (Aghabozorgi et al., 2014; Cen et al., 2015; Chatti et al., 2017; Demchenko et al., 2017; Klašnja-Milićević et al., 2017; Macfadyen et al., 2014; Sedkaoui, 2018), but the question remains on what should they do or learn to make this happen. For this reason, new stakeholders with respective roles have to be involved in this process to ensure that it meets technical and societal needs for working with BOLED analytics and enables to share knowledge and skills among them.

Since BOLED analytics is represented by the data lifecycle that consists of subsequent phases and each phase has its own characteristics and activities that need to be performed, most of the authors distinguished the roles according to respective phases (Lněnička & Komárková, 2019). Van den Broek, van Veenstra, and Folmer (2011) assigned five internal stakeholder roles to the various steps of the lifecycle: top management, information manager, legal advisor, community manager, and data owner. Attard, Orlandi, and Auer (2016) identified six roles in which stakeholders can participate to create value: data producer, data enhancer, data publisher, service creator, facilitator, and data consumer. Finally, Charalabidis, Loukis, and Alexopoulos (2014) and Chatti et al. (2017) reported that there is no more a clear distinction between providers and consumers of these data and everyone can produce and consume them as data prosumers.

From a methodological point of view, we follow the approach of Jongbloed et al. (2008) to deal with relevant stakeholders and their roles in the educational process. The approach consists of identifying stakeholders, classifying them according to their relative importance, and establishing working relationships with stakeholders. In order to distinguish between current roles in the educational process and emerging roles driven by BOLED analytics, we take into account the concept of Gonzalez-Zapata and Heeks (2015) who reported that there are two main groups of stakeholders' roles: primary (directly involved) and secondary (involved due to their knowledge and skills needed).

Table 3 provides a list of current roles of stakeholders in the educational process and their objectives for using BOLED analytics. The list extends the classification of Romero and Ventura (2010) who reported that there are different groups of involved stakeholders looking at educational information from different angles, according to their own mission, vision, and objectives. In order to achieve these objectives by respective roles, educational systems should find and support ways to gain relevant skills. Although the efforts are mostly focused on learners and educators, it is important to note that other roles of stakeholders should also profit from these changes, especially when some of the roles and objectives are overlapping.

The emerging roles are related to BOLED analytics. These roles are required for utilization of BOLED analytics. A certain classification of them was introduced by Lněnička and Komárková (2019). We put it into the educational context, aiming to formalize their relationships with current roles, see Fig. 4 and Fig. 5. The importance of these roles lies in providing additional layer that can be mapped onto current roles and build relationships serving for generating value from BOLED. Each of these emerging roles participates in different phases of the BOLED analytics lifecycle and ensures the continuity of each phase by means of activities and required outcomes.

Table 3 A list of current roles of stakeholders in the educational process

Current roles	Objectives for using BOLED analytics
Learners (students)	It should help them in personalizing and customizing their learning efforts through a recommendation of relevant activities, resources and learning tasks, either individually or in comparison to other students. Learners should be able to analyse their progress quickly, easily find new information and data sources, link them, suggest new learning paths, etc.
Educators (teachers and tutors)	It should help them to analyse learners' learning and behaviour including patterns, to predict learners' performance and classify them into groups, to publish data through various channels and enable learners to participate and create own learning content, to improve the adaptation and customization of courses and materials, etc.
Educational researchers	It should help them in exploring large amounts of data, constructing analytical models, comparing the efficiency of various techniques and recommending the most useful one for each task, developing specific data analytics tools for educational purposes, etc.
Course developers	It should help them to implement new platforms, tools, and services into courses as well as develop data-related features, enabled by robust data infrastructures. This also enables to evaluate the structure of course content and its effectiveness in the learning process. Course developers can also profit from the availability of data sources and findings of educational researchers.
Administrators (network, system, data)	It should help them to effectively organize and utilize available resources, to improve network and system efficiency and provide better services to users, to profit from benefits of new platforms, tools, and services, etc.
Educational institutions (secondary schools, universities)	It should help them in making decision enabled by greater availability of relevant information gained from analytics activities at a broader level. Linking big data from diverse sources enables to examine the structure and performance of a curriculum as a whole and focus more on learners and their needs in the educational process.
Education policy makers	It should provide them new insights into the educational process and examine if the learning outputs meet the needs of current educational developments. They should be able to implement them by defining the principles and government policies in the educational sphere regarding the needs of a data-driven education.

Skills required for performing BOLED analytics

As stated above, gaining value from these data also requires putting emphasis on developing new skills (Atenas et al., 2015; Yu & Wu, 2015). According to Mikroyannidis et al. (2016), organizations "*need their employees to understand the legal and economic aspects of data-driven business development, as a prerequisite for the creation of product and services that turn open and corporate data assets into decision-making insight and commercial value.*" As reported by Demchenko et al. (2017) and Gupta et al. (2015), a commonly accepted and harmonized instructional model that reflects all multidisciplinary knowledge and competencies is required. The educational model and approach have to solve different aspects and overlaps between topics that include both theoretical knowledge and practical skills. Regarding these changes, traditional roles such as teacher and student are evolving towards new roles that require additional skillsets. In addition, it is proper to investigate who should be responsible of shaping people with suitable skills to cope with these change (Coccoli et al., 2017). Chatti et al. (2017) and Khriyenko and Khriyenko (2013) claimed that this new environment is not limited to the particular institution, but it is driven the concept of data openness in which stakeholders are reusing data in an active way through cooperation, collaboration, and participation.

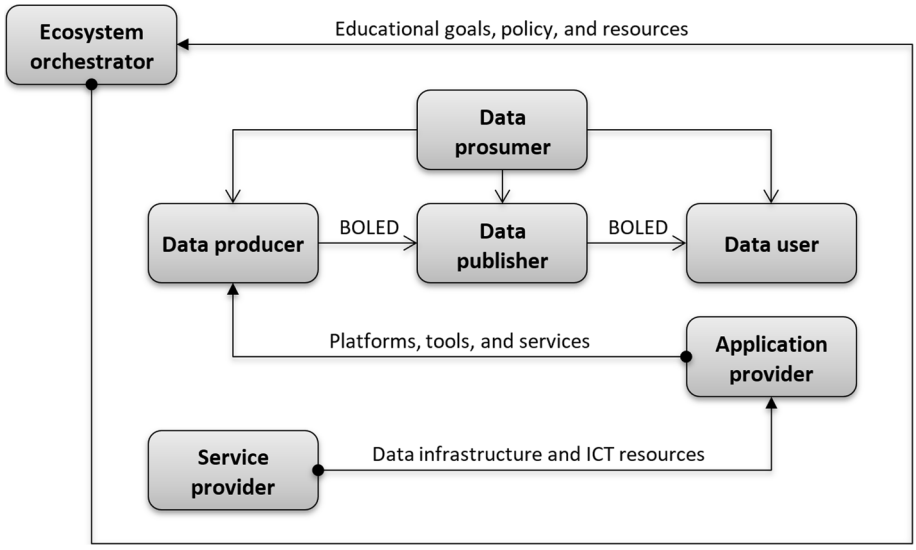

Fig. 4 The emerging roles in the BOLED analytics ecosystem. Source: Lněnička and Komárková (2019)

Thus, it is needed to identify involved stakeholders and their roles in the BOLED analytics ecosystem. Jetzek, Avital, and Bjorn-Andersen (2014) described data skills as "*the collective ability of individuals and organizations to use and reuse data*" and focus on access to data and data literacy. It is a multi-dimensional construct that encompasses equitable access opportunities as well as affordability of web access and data literacy of stakeholders. Missing insights on relationships, influences and dependencies during these interactions may affect the quality of the educational process and its continuous improvement (Cen et al., 2015; Demchenko et al., 2017; Rowley, 1997).

The findings pointed to a whole range of required skills required to be actively engaged in the BOLED analytics ecosystem and gain from its benefits. It is mainly due to the fact that the ecosystem is comprised of different stakeholders with diverging interests or preferences as well as approaches to deal with each phase and related activities of the BOLED analytics lifecycle. More precisely, Demchenko et al. (2017) reported that

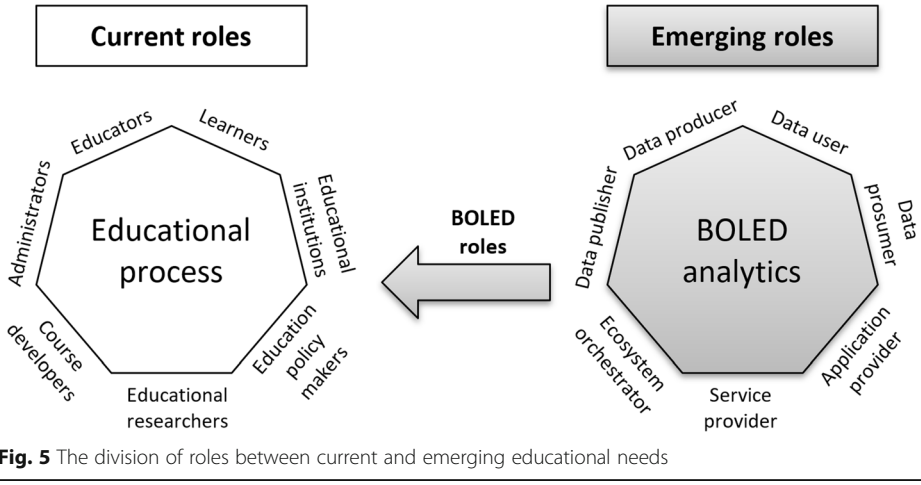

Fig. 5 The division of roles between current and emerging educational needs

"*modern data driven research and industry require new types of specialists that are capable to support all stages of the data lifecycle from data production and input to data processing and actionable results delivery, visualisation and reporting*". Therefore, there is no consensus framework on how these skills should be classified and what the required level is for different stakeholders.

The first group of authors are oriented toward emphasis on mastering hard skills, arguing that the real value lies in applying them on these data using different analytics techniques and tools (Demchenko et al., 2014; Li & Ni, 2015; Mikalef et al., 2018; Williamson, 2016; Zadeh et al., 2018). Lists of core areas required for common body of knowledge were introduced by Demchenko et al. (2014) or Mikalef et al. (2018). Among other, they include an emphasis on data management and challenges, security, anonymity, privacy and ethics of data, computing models, visualization and presentation of results, etc. Grillenberger and Romeike (2014) emphasised broader aspects of data management, e.g. still data storage (as in databases), but in combination with data usage and data analysis. Mason, Khan, and Smith (2016) proposed a conceptual realignment of the foundation skills of education to include being discriminate (discerning) alongside being literate and numerate. They emphasized the importance of data literacy.

However, focusing solely on data-specific jobs is not broad enough, a domain-specific knowledge is needed to span the respective sector (Aikat et al., 2017; Mikalef et al., 2018; Miller, 2014). Zadeh et al. (2018) focused on data analytics tools that facilitate learners' understanding of analytical concepts and practice. They highlighted the importance of domain knowledge and competency, technology and analytics skills, and critical and creative thinking. In addition, new roles such as educational data scientists, experts and programmers doing the data work, constructing the database architectures and designing the analytics will be necessary to fulfil these requirements (Mikalef et al., 2018; Williamson, 2016). Although Sedkaoui (2018) reported that "*creating value from available data requires a range of skills throughout the data lifecycle*", none of the authors attempted to map these requirements onto the all phases of the BOLED analytics lifecycle. The omission of the related activities may lead to bias in some types of analyses or incorrect interpretations and irreproducible results.

The other group of authors sees a need for acquiring both hard and soft skills. A skills perspective on how successful data reusers create value out of the available data sources was explored by Eckartz et al. (2016). They used a framework of IT, organization and skills required to innovate with data in which hard skills can be defined as more technical skills such as programming or data analytics skills and soft skills are more non-technical such as interdisciplinary cooperation or communication. These soft skills such as creativity and curiosity are needed to assess the possibilities of a dataset or which insights can possible be created (Atenas et al., 2015). According to Atenas, and Havemann, L (2015), the use of these data as OER require the development of critical, analytical, collaborative, and citizenship skills as well as learning good practices in data management, analysis, and reporting. Similarly, new adaptive forms of leadership, collaboration, policy development and strategic planning together with data access and security, data privacy and ethical dilemmas are required in case of successful institutional change for BOLED analytics implementation (Macfadyen et al., 2014; Mikroyannidis et al., 2016). A critical thinking about big data problems having appropriate computational skills and using appropriate technologies is emphasized by Sedkaoui (2018) and Song and Zhu (2015). Their

data ecosystem contains the following disciplines: leadership skills, communications skills, an eye for business and business values, project management skills, systems-thinking skills, big data technologies and the solution space, big data analytics lifecycle, and data management skills. Atenas et al. (2015) explored this issue in the context of transversal skills that include digital and data literacies, alongside skills for critical thinking, research, teamwork, and global citizenship. According to Aikat et al. (2017), training in teaming and leadership skills is a crucial component of data-centred learning. The findings of Huda et al. (2017) revealed that process and management skills should be engaged into competencies. It included commitment in planning, time management, and technology skills. In addition, the ability to use digital devices efficiently is an important part of an integrated skillset.

Another approach is given by Atenas, and Havemann, L (2015), who reported that reusing of these data should aim to help learners to develop a variety of skills, including digital and data literacy skills, research methods, problem solving and citizenship skills, i.e. the benefits for student engagement and participation in activities with real world relevance. Among the other benefits, Janssen et al. (2012) emphasized more participation and self-empowerment of users as well as stimulation of knowledge developments. Dubey and Gunasekaran (2015) distinguished between formal and non-formal education to hard and soft skills development. Finally, Coccoli et al. (2017) argued that the cooperation between academy and industry, which should learn from each other, would help achieving such a goal. They classified the required skills for working with BOLED into multi-disciplinary, multi-domain, multi-empathic, and multi-channel. In fact, people like these should have acquired degrees from both technical faculties and social sciences ones, which can give a technician the additional competences in teaching, social behaviour and interaction, communication, problem posing and solving, teamwork, creativity, and resilience (Coccoli et al., 2017).

Our framework aims to identify the necessary skills required for working with BOLED. The identification of skills arises from the literature review and the classification of stakeholders' current and emerging roles. The validity of the list of required skills is ensured by utilizing the Delphi method and establishing the expert opinion poll. The list is divided into three categories: hard and soft skills (general skills representing the well-known concepts among stakeholders), and BOLED analytics skills (data-related skills that need to be developed together with general skills).

Hard skills

By means of categorizing the required skills, the expert panel concurred that hard skills are the priority for stakeholders and distinguished them into core skills and domain and educational knowledge.

1.1. Core skills (foundational literacies):
   - 1.1.1. Data and information literacy – ability to read, understand, create and communicate data as information and effectively use that information for an intended purpose.
   - 1.1.2. Numeracy – mathematics and statistics knowledge that deals with the collection, analysis, and interpretation of numerical data.
   - 1.1.3. ICT and scientific literacy – a set of abilities needed to accomplish engineering, scientific or computer-related tasks.

1.1.4. Financial literacy – ability to understand the current and future value of data.

1.2. Domain and educational knowledge:

1.2.1. Knowledge of the sector – understanding data sources and real-world situations and processes behind the existing data, including data security and privacy.

1.2.2. Knowledge of the topic – awareness of educational goals and processes, knowing the educational questions that matter and how data fit with the topic.

Soft skills

Reaching a consensus on the list of soft skills took more time. These skills were distinguished into competencies and character qualities.

1.3. Competencies are a set of abilities required to act and react when dealing with routine tasks as well as new tasks.

1.3.1. Collaboration and team-working – ability to work with multiple stakeholders from different disciplines in order to achieve or do something.

1.3.2. Communication – ability to reach understanding and exchange information using available communication tools.

1.3.3. Critical thinking and problem solving – ability to critically work through complex problem, identify its key components, and find a solution or solutions.

1.3.4. Analytical thinking – ability to re-formulate and decompose complex problem to trace the cause and implications in a systematic manner.

1.3.5. Conceptual thinking – ability to identify patterns or relationships between situations that are not obviously related.

1.3.6. Storytelling – ability to transform ideas and insights into actionable and well communicated recommendations and reports.

1.3.7. Lifelong learning – ability to acquire new qualifications constantly, be flexible and adapt on changing situations and draw conclusions from mistakes.

1.4. Character qualities are abilities beyond academic learning, knowledge and skills.

1.4.1. Motivation and initiative – ability to direct attention and effort in accomplishing the challenging tasks.

1.4.2. Creativity and curiosity – ability to generate ideas, new solutions to problems, and explore data from different.

1.4.3. Adaptability and flexibility – ability to adapt and work effectively in various and changing situations and activities.

1.4.4. Achievement orientation and proactivity – ability to concern for working towards a standard of excellence.

1.4.5. Persistence and coping with stress – ability to concentrate and focus on the aim despite difficulties and discouragement.

1.4.6. Responsibility – ability to face up consequences of behaviours and actions.

1.4.7. Leadership – ability to influence people toward the achievement of objectives.

1.4.8. Self-confidence – ability to rely on the self-knowledge and one's own skills.

1.4.9. Social and cultural awareness – ability to be able to work with data in the context of social and cultural background and changes that may affect the outputs.

BOLED analytics skills

Due to specific requirements and approaches needed to work with BOLED, it was agreed that new category of skills representing the BOLED analytics lifecycle phases and activities has to be added to a set of required skills. The set of skills for working with BOLED should comprise the skills needed to transform raw and mostly unstructured data into value-added, fact-based insights, with an emphasis on relevant platforms, tools, and services occurring in each phase of the data lifecycle, see Fig. 6. The BOLED analytics lifecycle was introduced in Lněnička and Komárková (2019).

Since BOLED analytics is a complex process involving multiple activities that may or may not be necessarily performed to gain value from these data, a three-level skills matrix was proposed to better reflect required skills for different roles, see Fig. 7. It was firstly developed by Lněnička et al. (2018). The first level is represented by general skills, i.e. hard and soft skills. Hard skills are further divided into core skills (foundational literacies) and domain and educational knowledge while soft skills can be classified into competencies and character qualities. On the second level, scenario-based skills related to each phase and activities of the BOLED analytics lifecycle are defined. Both these levels were rated for different roles of stakeholders representing a third dimension. We note that the difference between hard skills and BOLED analytics is important due to their focus. While hard skills are general and can be utilized across many subjects and disciplines, BOLED analytics skills are specific to particular data lifecycle phases.

Results

The skills that involved stakeholders must have to become advanced in BOLED analytics are in exceptionally high demand. In this regard, the Delphi-based expert opinion poll was established to provide recommendations for educational institutions on how to improve their educational programs and research to better target critical skills gaps. More precisely, the development of more efficient educational processes is in great demand to help learners and other stakeholders with acquiring the skills to gain value from available data. Our results provide the set of skills classified into three categories: hard, soft, and BOLED analytics skills. Each of them was evaluated by experts on a five-point Likert scale, from 1 = extremely unimportant to 5 = extremely important. The importance of each skill was quantified for each role that stakeholders play in and around the educational process. Both categories of current and emerging roles were evaluated to show various demands and needs of involved stakeholders. However, it should be noted that current roles are primary in this study and emerging roles are representing the link to necessary skills that should be absorbed by current roles.

The results are displayed in Fig. 8 and Fig. 9 as mean values and standard deviations respectively. The mean value for the set of skills is highest for educational researchers (4.83) and lowest for service provider (3.92). In the case of the average standard deviation, it is highest for learners (0.63) and lowest for educational researchers (0.18). The range of opinions on the desired set of skills for learners was highest in the case of hard skills because experts expressed their worries about the constantly changing nature of jobs. Furthermore, it can be argued what is the level of hard skills needed for educational institutions and education policy makers.

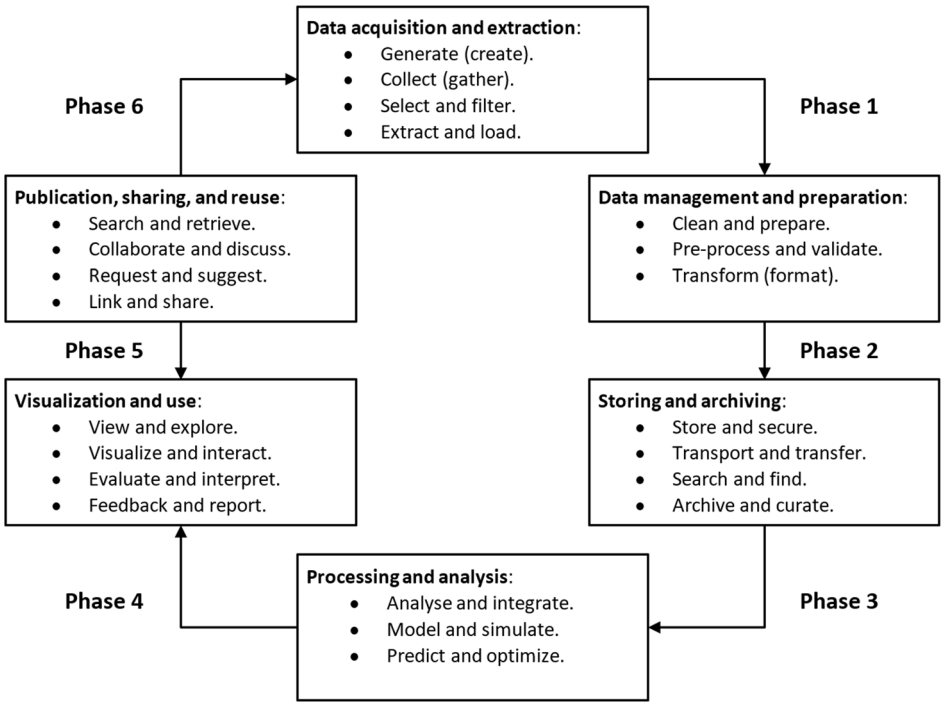

Fig. 6 The BOLED analytics lifecycle phases. Source: Lněnička and Komárková (2019)

Based on the stakeholder theory, Table 4 represents the multiple and interdependent interactions that simultaneously exist in stakeholder environments and serve to predict organizational responses to these needs. Figure 10 conceptualizes the interactions between current roles and appropriate emerging roles from Table 4. These differentiate by types of lines. The figure also includes the values of hard, soft, and BOLED analytics skills from Fig. 8. The skills for the application provider role are key for most of the current roles, because they consist of not only delivering applications, but especially include knowledge on how to use relevant platforms, tools, and services, and in which data lifecycle phases should be correctly deployed and used. Further required skills are related to roles in which stakeholders are working with BOLED.

According to the results, hard skills are key for educators, educational researchers, and data producers. These roles are closely related to teaching and managing educational resources with strong emphasis on ability to understand relevant data sources, interpret their value and establish educational processes to engage other stakeholders and transfer knowledge. Soft skills should be a priority for educators, educational researchers, ecosystem orchestrators, and data prosumers. Their importance lies in establishing a set of conditions necessary for the learning process, including enabling communication, cooperation, negotiation, and information sharing among the stakeholders. Skills related to BOLED analytics are important for educational researchers, data producers, and data prosumers. The phase that requires soft skills the most is data publication, sharing, and reuse since stakeholders of all roles should participate here and bring together their skills to create value.

The expert panel agreed on data and information literacy as the most important skill and argued that it should be a critical part of any role. Regarding softs skills,

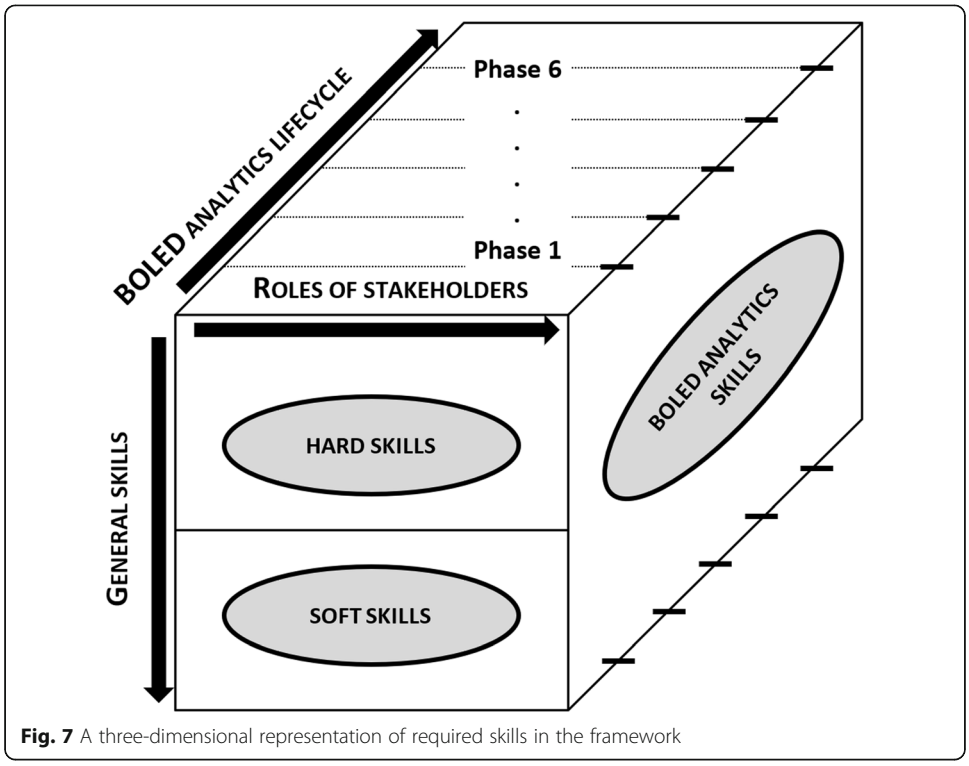

Fig. 7 A three-dimensional representation of required skills in the framework

adaptability and flexibility are considered crucial for working with BOLED. All the stakeholders should focus on these BOLED activities, which were rated as high priority: (1) view and explore, (2) feedback and report, and (3) search and find. Although being able to view and explore data sets visually is the most important activity, visual information is not always accessible through data analytics tools or open data portals and the potential to engage more stakeholders is not fully used. Therefore, on the one hand,

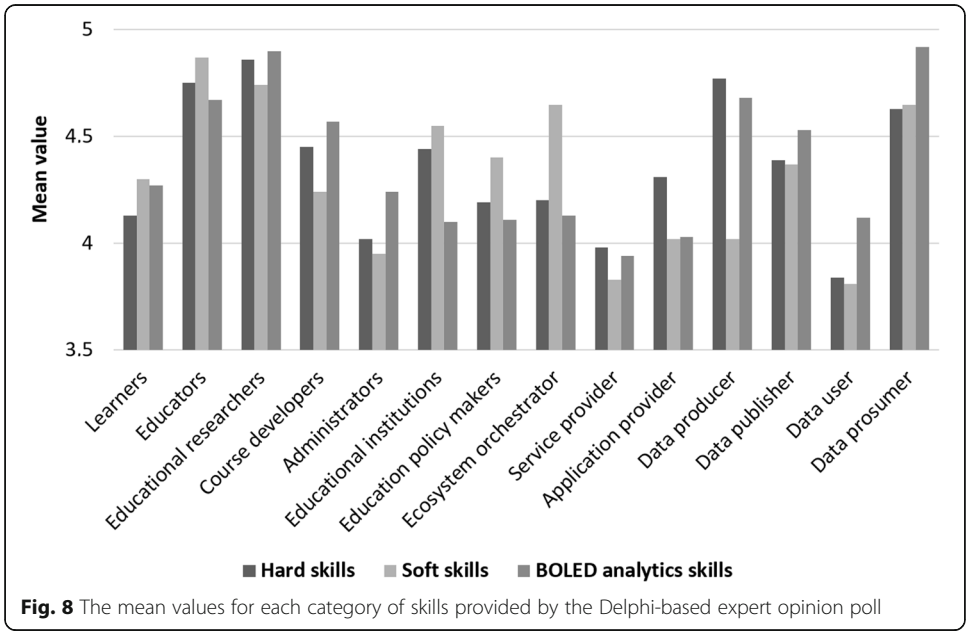

Fig. 8 The mean values for each category of skills provided by the Delphi-based expert opinion poll

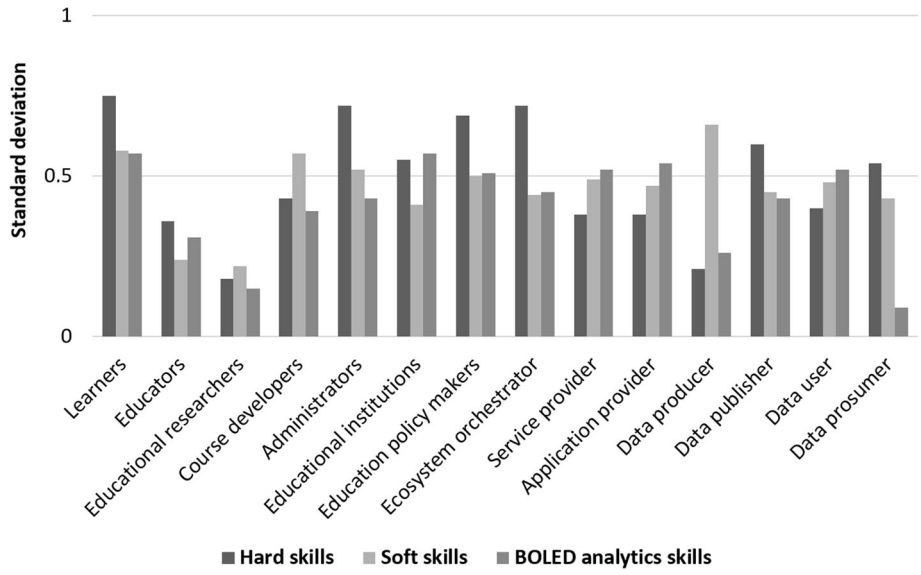

Fig. 9 The mean standard deviation values for each category of skills provided by the Delphi-based expert opinion poll

it is necessary to increase emphasis on data visualization techniques in the educational process and, on the other hand, tools developers and data providers should introduce more features to enable applying these skills in practice. It should be also noted that multiple sets of skills, in different proportion, can be required by more than one role and those roles may be overlapping. In this regard, not all the BOLED analytics lifecycle phases need to be performed by one role and it is possible and recommended to involve more stakeholders in different roles to work together. For this purpose, soft skills related to communication, collaboration, and cooperation are important to overcome the missing hard skills by some stakeholders.

The prioritization of current and emerging roles is important for assessing their impact on planning and implementation of suitable changes in the educational process regarding BOLED analytics. Each identified role is mapped onto an influence-impact grid to show the level of active involvement a role has to effect changes the development and implementation of BOLED analytics in the educational process. These are

Table 4 Mapping current roles onto their roles in the BOLED analytics ecosystem

	Ecosystem orchestrator	Service provider	Application provider	Data producer	Data publisher	Data user	Data prosumer
Learners	NO	NO	YES	YES	NO	YES	YES
Educators	NO	NO	YES	YES	YES	NO	YES
Educational researchers	NO	NO	YES	YES	YES	YES	YES
Course developers	NO	NO	YES	NO	YES	NO	YES
Administrators	NO	YES	YES	NO	NO	NO	NO
Educational institutions	YES	YES	NO	YES	YES	NO	NO
Education policy makers	YES	NO	NO	NO	NO	NO	NO

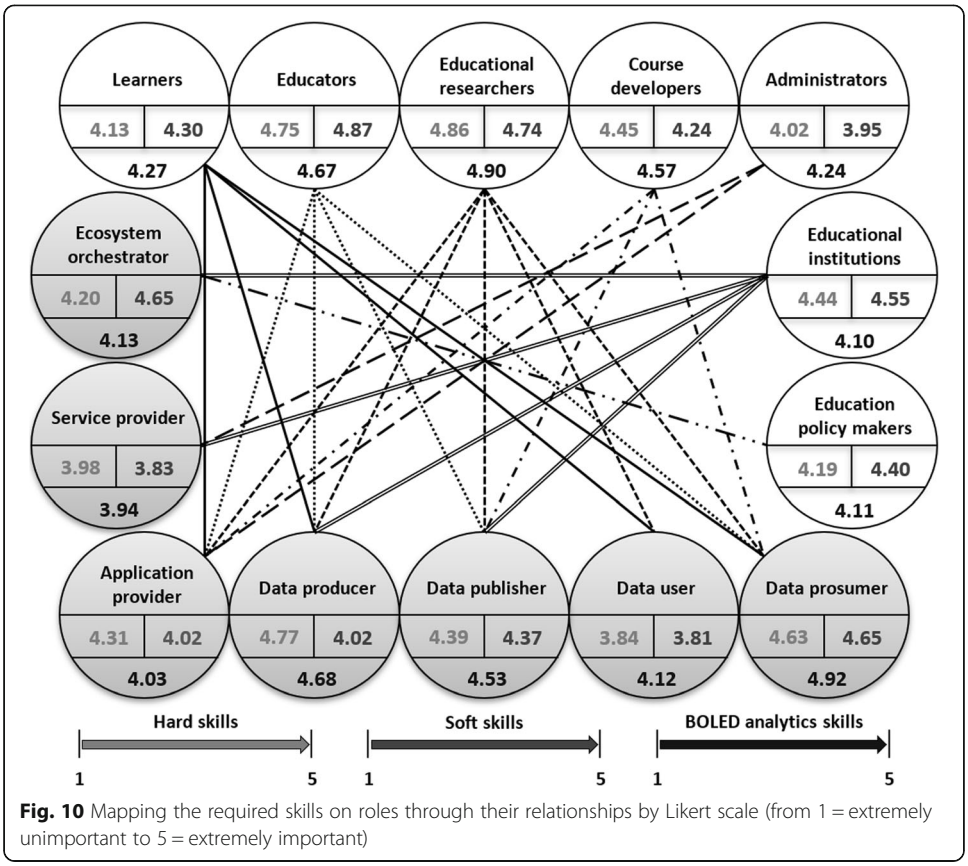

Fig. 10 Mapping the required skills on roles through their relationships by Likert scale (from 1 = extremely unimportant to 5 = extremely important)

displayed in Fig. 11. The prioritization should help organizations and institutions in higher education having decision-making authority in responding to relevant stakeholders' pressures. The roles that should be managed closely are educational researchers, education policy makers, and ecosystem orchestrators. Researchers explore new opportunities, create impulses, and transfer them in close cooperation with educators into the educational process. For policy makers, it is important that they understand the value of BOLED analytics and support the development of appropriate skills, especially regarding financial resources that flow through the ecosystem. The responsibility of ecosystem orchestrators lies in defining and managing the ecosystem by regulation and enforcement of required activities and operational guidelines in the context of educational goals, policy, and resources. The roles that should be satisfied are mostly customers' roles and should be managed accordingly. Other roles should always be informed to show them that they are clearly considered as a part of the ecosystem.

Discussion

Our findings are in line with findings of Aikat et al. (2017), Coccoli et al. (2017), Eckartz et al. (2016) and Ferguson et al. (2016), who recommend to engage more stakeholders and set up multi-disciplinary teams focused on flexibility and interdisciplinary competences. This requires the need of some kind of orchestration, rather than highly specific skills (Coccoli et al., 2017). Aikat et al. (2017) reported that it must incorporate interdisciplinary training that couples the domain sciences with data science. When the teams are successful, it may motivate and engage more teams and help to create inter-organizational networks with complementing skills and experience

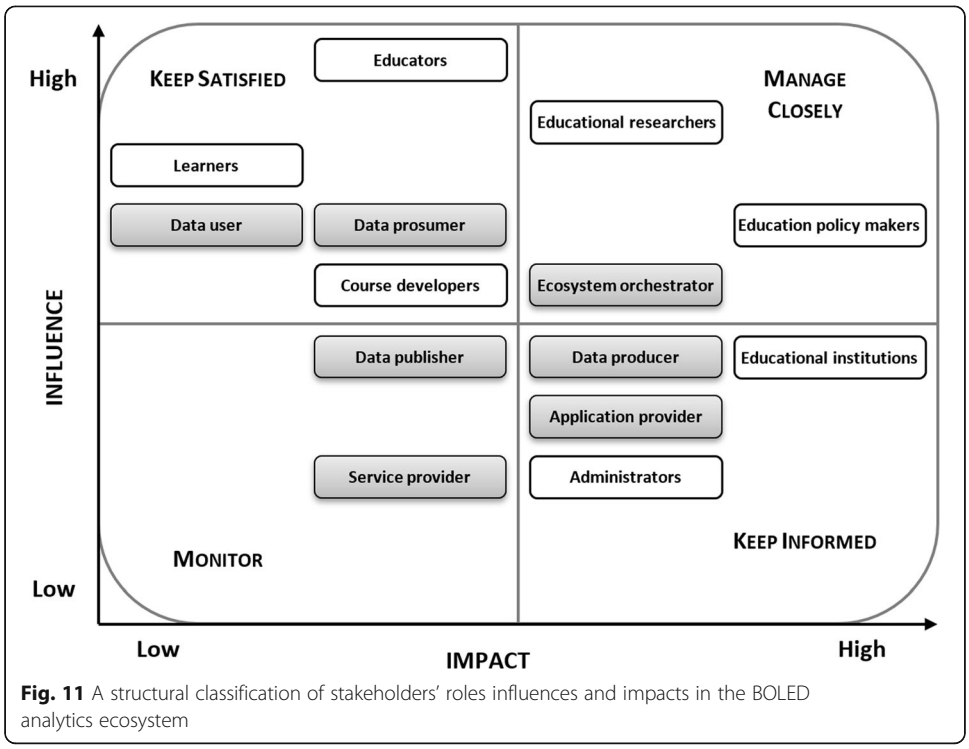

Fig. 11 A structural classification of stakeholders' roles influences and impacts in the BOLED analytics ecosystem

(Eckartz et al., 2016). The need of skills related to inter-disciplinary team management was also discussed by Mikalef et al. (2018). In addition to the findings outlined above, Atenas, and Havemann, L (2015) argued that skills should also include an understanding of the laws in relation to these data including their interpretation and limits, together with more detailed explanation, guidance and clarification. Ferguson et al. (2016) stated that this will require a focus on the skills required in different areas, the provision of support for educational institutions, educators and other educational staff. It is also expected that organizations will be involved in this process, secure new talents and train their existing staff into becoming proficient data practitioners (Mikroyannidis et al., 2016).

In order to enable communication, cooperation or team-working between stakeholders in different roles, integrated, connected and technology-supported learning environments and educational labs have to be available for all the involved stakeholders. There is a growing number of free tools and software packages easily accessible online with which data analytics becomes much easier, even without a technical background (Lněnička & Komárková, 2015). This enables encouraging more stakeholders to try some of the BOLED analytics activities and gain an insight into the processes behind these data. Among the platforms, tools, and services that should be used in the educational process to facilitate understanding of analytical concepts and practice we recommend to begin with open data portals that provide features to easily search, filter, analyse, link, and visualize open data sets. More advanced features are offered by user interface based tools such Rapid Miner, Tableau, Weka etc. Robust desktop tools for BOLED analytics are represented by Apache Hadoop, Apache Spark, Apache Marmotta etc. While existing educational programs should be modified with accordance to sets of skills identified in our study, these skills may also be acquired through external OER,

MOOC and other online resources. Regarding the specific industry needs, educational programmes should be also tailored accordingly.

A second issue to discuss is data quality that may affect the skills required to utilize BOLED analytics. Other data-related problems are may include task complexity, e.g. lack of ability to discover the appropriate data or no explanation of the meaning of data (Janssen et al., 2012), lack of interoperability of institutional data systems and the complexity of managing and analysing large amount of heterogeneous data, the loss of value and / or reliability of data coming from different systems (Abdelouarit et al., 2015; Zuiderwijk et al., 2012). Pecori (2018) addressed issues related to technologies integration, user acceptability, and the security perspective, such as ownership and privacy of data. Insufficient awareness of educators and other stakeholders in data security is another issue that may need to be addressed. Among risks that may influence the success of BOLED analytics, the weak or missing institutional support from the educational institutions and education policy makers is the most important one (Janssen et al., 2012; Macfadyen et al., 2014). Lack of acceptance among stakeholders may hinder the implementation of these approaches into the educational process (Lněnička et al., 2018).

Another point to take into the account, when working with BOLED, is considering investment in security (Sedkaoui, 2018). The question of security and privacy in the educational context is widely discussed among the researchers (Aguilar, 2018; Chatti et al., 2017; Demchenko et al., 2014; Klašnja-Milićević et al., 2017; Kyritsi et al., 2019; Macfadyen et al., 2014; Pecori, 2018). It is especially challenging to apply appropriate anonymization techniques for releasing data sets without compromising personal privacy. On the other hand, this approach raises technical problems such as the loss of a certain amount of information found in the original data (Kyritsi et al., 2019). Thus, Educational institutions should assess the value that their data hold and then deal with privacy infringement, publication of data against the law and improper or inaccurate data, misinterpretation of data, etc. (Atenas, & Havemann, L. , 2015). Miller (2014) recommended to set minimum standards for data and analytics literacy required by all learners in the age of big data. These literacy training should be created and delivered via MOOC. In addition, open online communities should be established to engage industry, government, and academia due their shared interests and importance of open data portals (Lněnička & Máchová, 2015). Finally, establishing working groups to address key data policy issues such as information security, individual privacy, and the ethical use of BOLED is important to ensure that the value from these data will be gained (Miller, 2014; Song & Zhu, 2015).

Furthermore, it is essential that data users and suppliers discuss their data needs and ensure that data are more reliable, available and usable for external stakeholders, i.e., the supply of skills matches demand (Dubey & Gunasekaran, 2015; Eckartz et al., 2016). In addition, data sharing and reuse in the educational field needs further research to explore whether the context and scope of the dataset collection can significantly affect its potential reuse (Verbert et al., 2012), especially in the context of visual learning analytics and how these data are presented to different stakeholders (Vieira et al., 2018). Since the opportunities presented by BOLED analytics are only beginning to emerge (Aikat et al., 2017) and they are changing in very fast cycles (Mikalef et al., 2018), existing skills may become obsolete very fast and the exact set of skills may be hard to identify. In this regard, Eldridge et al. (2018) explored the possibility of replacing humans and their

skills with software solutions and concluded that it is important to facilitate complementarity and obtaining the best contribution from both.

A third issue is related to the fact that since different people have modes of thinking, levels of knowledge, and ability of learning, their learning effectiveness and efficiency can be different even with exactly the same learning conditions and environment (Cen et al., 2015). As stated by Attard et al. (2016), the motivations, levels of expertise, and priorities of stakeholders differ. Therefore, these levels have to be considered to address both literacy and specialized skills at all levels from undergraduate to executive education (Miller, 2014). The digital divide will alienate many data consumers who are unable to acquire or employ the technical skills to access or decipher these data (Millette & Hosein, 2016). Khriyenko and Khriyenko (2013) suggested that on top of the basic level there should be provided courses with a high level of flexibility from to accentuate the stakeholder's best abilities and skills, assess the stakeholder's potential and develop it further. Dubey and Gunasekaran (2015) identified training as a moderating variable and controlled for the demographic profiles of the learners to further account for differences in learning ability. On the other hand, a conservative approach of educators and their unwillingness to use new technologies may affect the implementation of these approaches.

Recommendations, limitations and future research

This research assumes that there is an interest of educational institutions in benefits provided by BOLED analytics. As reported by Mikalef et al. (2018), while some of the relevant competencies are indeed developed while working, a large part of the essential skills should be developed in higher education. Hence, our findings aimed toward helping educational institutions, especially higher educational institutions, to reorient themselves to a world where data analytics is at the core of all.

Based on our results, we formulated the following recommendations in relation to objectives of stakeholders' current roles in higher education institutions. **Learners** should be taught appropriate skills that are essential for working with BOLED. First, it is necessary to have a good understanding of the problem and topic to be able to answer the right questions. Before giving examples from practice it is necessary to provide the theoretical background for this topic. After that, learners should be given some practice exercises or activities on how to get significant value from BOLED. More precisely, they should focus on developing BOLED analytics skills relevant to data acquisition and extraction, visualization and use, and tasks encompassing data publication, sharing, and reuse. They need to be able to find the right data sets at the first place and then collect, select and filter them. Finally, it is expected that students will engage in industry practicums and internships to train for real-world application of skills.

The skills base of **educators** needs to be extended to include an understanding of analytical techniques relevant to phases of BOLED analytics lifecycle. They need to acquire new skills to perform analytics on these data, especially at the beginning of the process, where they have to target the right data sets and transform them into formats needed for the further analysis. It relates to pedagogical content knowledge and the responsibility to provide learners knowledge and skills needed to make data-driven decisions. They should be also able to handle tasks related to data management, storing, processing, and visualization and use. In this role, the soft skills (good communication, storytelling, critical and analytical thinking) are just as important as the hard skills.

Educational researchers need to be equipped with BOLED analytics skills to link, analyse, and interpret complex patterns of data along with the knowledge on how to utilize these findings in the educational process. This role should include soft skills relevant to critical thinking and problem solving, analytical thinking, and conceptual thinking. It also requires motivated and initiative-oriented as well as creative and curious stakeholders to generate novel solutions to problems based on these data. The storytelling is another important part of exploring data. In order to offer the results that will be useful and applicable, educational researchers have to engage in social interactions and deal with involved stakeholders from a diverse background.

For the role of a **course developer**, which is often overlapping with the roles of educators and educational researchers, domain and educational knowledge is important together with the ability to transfer knowledge to other stakeholders through modern ICT and data analytics tools, platforms, and services. Our research shows that the primary skills for **administrators** are in the domain of data and infrastructure management. This role is more technical than others and requires a good understanding of processes behind BOLED analytics, especially regarding ICT resources. One of the core characteristics of **educational institutions** is the ability to understand the function of the domain areas in which BOLED analytics is going to be utilized. Not only educational knowledge is required, but knowing what jobs are in high demand is also crucial. This role functions as an intermediary between other stakeholders. **Education policy makers** must have the skills to communicate. It is very important to be able to discuss with other stakeholders, understand the problems they are facing, and provide a solution that will be satisfactory. Both these roles highly depend on soft skills and social competences.

Other recommendations towards improving curricula and teaching models to better adapt to these educational challenges should be focused on systematically developing and strengthening both hard and soft skills as well as scenario-based BOLED analytics skills. Courses, practices and activities should be designed to enable participation, collaboration and cooperation between stakeholders, especially field practitioners, specialists and experts. The pedagogical approach should be revised in line with these findings. The skills development process should integrate all the skills identified and evaluated in this study. Practical examples and best practices using relevant platforms, tools, and services should be also incorporated into educational process. In this regard, stakeholders should be able to choose the right one for each phase or activity of the BOLED analytics lifecycle. Steps of developing models or decision matrices for this issue should be also a part of courses. Stakeholders must also acquire skills to use and dialog with different digital devices and interfaces. With the emergence of ambient intelligence and the abundance of smart and IoT-enabled devices, the need for these skills will increase. Finally, the structural classification of influences and impacts of particular roles in the BOLED analytics ecosystem is provided to prioritize further steps regarding involved stakeholders.

However, there are also limitations that can affect results obtained in this study. The number of experts may be seen as insufficient in terms of the Delphi process and the sample may not have been fully representative. The reliability may be questioned since the existence of personal and situation-specific bias means that every new application of the method involves the creation of a new measuring instrument (Grisham, 2009;

Okoli & Pawlowski, 2004). The risk of bias associated with the Delphi method was dealt with in the following ways: 1) the selection process of the experts' panel members was especially designed to avoid drop-out of experts and ensure the incorporation of different experts' from a wide range of fields who are qualified to answer the questions; 2) all the components serving as inputs to our study resulted from a formal process that included the comprehensive literature review and concepts from the stakeholder theory in higher education and its communities. Our results indicate that under the appropriate circumstances, Delphi is a useful consensus method since it inherently provides richer data because of their multiple iterations and their response revision due to feedback (Okoli & Pawlowski, 2004).

Finally, it should be noted that this research provides a general view on the issue of skills needed to perform BOLED analytics. Since the education of professionals in different fields requires different skillsets and approaches, future research should be focused on sector-specific skills and their development regarding curricula at educational institutions.

Conclusions

In the era of the data-driven economy in which data-driven decisions are critical for organizations, abilities to work with different data sources in order to turn them into knowledge should be included in the educational process. Whereas literature has mainly focussed on the role that ICT can play in facilitating learners and other stakeholders in this process, there is a significant need to understand and define required skills to work and reuse these data. Hence, the increased application of data analytics requires a new generation of experts with unique interdisciplinary competences. In this regard, it is challenging to identify the type of skills that need to be provided and find the proper ways to develop them. In contrast, many traditional skills will be reshaped due to new features of BOLED analytics. This paper aims to stimulate higher education institutions to develop new set of skills that are intended to transform these data into value.

Since the problem of required skills can benefit from subjective judgments on a collective basis, a Delphi method is used to measure the diversity of opinions on this topic that spans a wide range of disciplines from a panel of experts. Our study shows that BOLED analytics requires the development of interdisciplinary competencies that will include hard skills related to data and information literacy together with the ability to use relevant tools, domain knowledge, and soft skills enabling to collaborate and communicate effectively, and be proactive and flexible to adapt to ongoing challenges. In addition, stakeholders should be also able to choose the right platforms, tools, and services for each phase of the BOLED analytics lifecycle.

We introduced sets of skills for both current and emerging roles that nowadays occur in the education ecosystem. We classified them into three categories: hard, soft, and BOLED analytics skills. Selected experts evaluated their importance for different roles in which learners, educators and other stakeholders work with these data. In order to provide a bridge between current roles in which stakeholders participate in the educational process and emerging roles needed to manage and unlock data potential, we mapped current roles onto respective emerging roles. The structural classification of

influences and impacts of particular roles in the BOLED analytics ecosystem enables to prioritize further implementation steps regarding working with involved stakeholders.

While governments have a fundamental role to play in providing the educational basis and educational institutions should deliver their courses in an efficient and responsive manner, the private sector should be able provide some training and sector-specific skills. This is also one of the limitations that can hinder the efforts because educational institutions need to know what learners' and other stakeholders' skills should be. The current level of skills of people entering the higher education system can also be limiting. Therefore, future research should be focused on sector-specific skills and required levels of skills for respective jobs.

Acknowledgements
Not applicable.

Additional information
This study is based on our earlier work presented in Lněnička et al. (2018).

Transparency note
This study is based on our earlier work presented in Lněnička et al. (2018).

Authors' contributions
Each author contributed equally to the manuscript. The authors read and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 11 November 2019 Accepted: 27 April 2020
Published online: 19 August 2020

References

Abdelouarit, K. A., Sbihi, B., & Aknin, N. (2015). Big-learn: Towards a tool based on big data to improve research in an e-learning environment. *International Journal of Advanced Computer Science and Applications*, *6*(10), 59–63.

Aghabozorgi, S., Mahroeian, H., Dutt, A., Wah, T. Y., & Herawan, T. (2014). An approachable analytical study on big educational data mining. In B. Murgante et al. (Eds.), *Computational science and its applications – ICCSA 2014*, (pp. 721–737). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09156-3_50.

Aguilar, S. J. (2018). Learning analytics: At the nexus of big data, digital innovation, and social justice in education. *TechTrends*, *62*(1), 37–45. https://doi.org/10.1007/s11528-017-0226-9.

Aikat, J., Carsey, T. M., Fecho, K., Jeffay, K., Krishnamurthy, A., Mucha, P. J., … Ahalt, S. C. (2017). Scientific training in the era of big data: A new pedagogy for graduate education. *Big Data*, *5*(1), 12–18. https://doi.org/10.1089/big.2016.0014.

Atenas, J., & Havemann, L. (2015). *Open data as open educational resources: Case studies of emerging practice*. London: Open Knowledge, Open Education Working Group. https://doi.org/10.6084/m9.figshare.1590031.

Atenas, J., Havemann, L., & Priego, E. (2015). Open data as open educational resources: Towards transversal skills and global citizenship. *Open Praxis*, *7*(4), 377–389. https://doi.org/10.5944/openpraxis.7.4.233.

Attard, J., Orlandi, F., & Auer, S. (2016). Data driven governments: Creating value through open government data. In A. Hameurlain et al. (Eds.), *Transactions on large-scale data- and knowledge-centered systems XXVII. Lecture notes in computer science, 9860*, (pp. 84–110). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-53416-8_6.

Cen, L., Ruta, D., & Ng, J. (2015). Big education: Opportunities for big data analytics. In *2015 IEEE international conference on digital signal processing (DSP)*, (pp. 502–506). IEEE. https://doi.org/10.1109/ICDSP.2015.7251923.

Cervone, H. F. (2016). Organizational considerations initiating a big data and analytics implementation. *Digital Library Perspectives*, *32*(3), 137–141. https://doi.org/10.1108/DLP-05-2016-0013.

Charalabidis, Y., Loukis, E., & Alexopoulos, C. (2014). Evaluating second generation open government data infrastructures using value models. In *Proceedings of the 47th Hawaii international conference on system sciences*, (pp. 2114–2126). IEEE. https://doi.org/10.1109/HICSS.2014.267.

Chatti, M. A., Muslim, A., & Schroeder, U. (2017). Toward an open learning analytics ecosystem. In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education*, (pp. 195–219). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5_12.

Coccoli, M., Maresca, P., & Stanganelli, L. (2017). The role of big data and cognitive computing in the learning process. *Journal of Visual Languages and Computing*, *38*, 97–103. https://doi.org/10.1016/j.jvlc.2016.03.002.

Colpaert, J. (2018). Exploration of affordances of open data for language learning and teaching. *Journal of Technology and Chinese Language Teaching*, *9*(1), 1–14.

Demchenko, Y., Belloum, A., de Laat, C., Loomis, C., Wiktorski, T., & Spekschoor, E. (2017). Customisable data science educational environment: From competences management and curriculum design to virtual labs on-demand. In *2017 IEEE International Conference on Cloud Computing Technology and Science*, (pp. 363–368). IEEE. https://doi.org/10.1109/CloudCom.2017.59.

Demchenko, Y., Gruengard, E., & Klous, S. (2014). Instructional model for building effective big data curricula for online and campus education. In *2014 IEEE 6th international conference on cloud computing technology and science*, (pp. 935–941). IEEE. https://doi.org/10.1109/CloudCom.2014.162.

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.

Eckartz, S., van den Broek, T., & Ooms, M. (2016). Open data innovation capabilities: Towards a framework of how to innovate with open data. In H. J. Scholl et al. (Eds.), *Electronic government: Proceedings of the 15th IFIP WG 8.5 international conference, EGOV 2016*, (pp. 47–60). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-44421-5_4.

Eldridge, C., Hobbs, C., & Moran, M. (2018). Fusing algorithms and analysts: Open-source intelligence in the age of 'big data'. *Intelligence and National Security*, *33*(3), 391–406. https://doi.org/10.1080/02684527.2017.1406677.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., … Vuorikari, R. (2016). *Research evidence on the use of learning analytics: Implications for education policy*. Seville: Joint Research Centre.

Fernández, A., Peralta, D., Benítez, J. M., & Herrera, F. (2014). E-learning and educational data mining in cloud computing: An overview. *International Journal of Learning Technology*, *9*(1), 25–52. https://doi.org/10.1504/IJLT.2014.062447.

Geiger, C. P., & von Lucke, J. (2012). Open government and (linked) (open) (government) (data). *JeDEM-eJournal of eDemocracy and Open Government*, *4*(2), 265–278. https://doi.org/10.29379/jedem.v4i2.143.

Gkontzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2019). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2019.1709209.

Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, *32*(4), 441–452. https://doi.org/10.1016/j.giq.2015.09.001.

Grillenberger, A., & Romeike, R. (2014). Big data – Challenges for computer science education. In Y. Gülbahar, & E. Karataş (Eds.), *International conference on informatics in schools: Situation, evolution, and perspectives*, (pp. 29–40). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09958-3_4.

Grisham, T. (2009). The Delphi technique: A method for testing complex and multifaceted topics. *International Journal of Managing Projects in Business*, *2*(1), 112–130. https://doi.org/10.1108/17538370910930545.

Gupta, B., Goul, M., & Dinter, B. (2015). Business intelligence and big data in higher education: Status of a multi-year model curriculum development effort for business school undergraduates, MS graduates, and MBAs. *Communications of the Association for Information Systems*, *36*(1), 449–476.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98–115. https://doi.org/10.1016/j.is.2014.07.006.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: Theory and technology*, *1*(1), 1–136.

Huda, M., Maseleno, A., Shahrill, M., Jasmi, K. A., Mustari, I., & Basiron, B. (2017). Exploring adaptive teaching competencies in big data era. *International Journal of Emerging Technologies in Learning (iJET)*, *12*(03), 68–83. https://doi.org/10.3991/ijet.v12i03.6434.

Iiyoshi, T., & Kumar, M. S. V. (Eds.) (2008). *Opening up education: The collective advancement of education through open technology, open content, and open knowledge*. Boston: MIT Press.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740.

Janssen, M., & Kuk, G. (2016). Big and open linked data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 3–13. https://doi.org/10.1080/10919392.2015.1124005.

Jena, R. K. (2019). Sentiment mining in a collaborative learning environment: Capitalising on big data. *Behaviour & Information Technology*, *38*(9), 986–1001. https://doi.org/10.1080/0144929X.2019.1625440.

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 100–120. https://doi.org/10.4067/S0718-18762014000200008.

Jongbloed, B., Enders, J., & Salerno, C. (2008). Higher education and its communities: Interconnections, interdependencies and a research agenda. *Higher Education*, *56*(3), 303–324. https://doi.org/10.1007/s10734-008-9128-2.

Khriyenko, O., & Khriyenko, T. (2013). Innovative education environment and open data initiative: Steps towards user-powered society-oriented systems. *GSTF Journal on Computing (JoC)*, *3*(3), 31–39. https://doi.org/10.7603/s40601-013-0020-2.

Klašnja-Milićević, A., Ivanović, M., & Budimac, Z. (2017). Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education*, *25*(6), 1066–1078. https://doi.org/10.1002/cae.21844.

Kyritsi, K. H., Zorkadis, V., Stavropoulos, E. C., & Verykios, V. S. (2019). The pursuit of patterns in educational data mining as a threat to student privacy. *Journal of Interactive Media in Education*, *2019*(1), 2. https://doi.org/10.5334/jime.502.

Li, S., & Ni, J. (2015). Evolution of big-data-enhanced higher education systems. In *2015 eighth international conference on internet computing for science and engineering*, (pp. 253–258). IEEE. https://doi.org/10.1109/ICICSE.2015.53.

Liñán, L. C., & Pérez, Á. A. J. (2015). Educational data mining and learning analytics: Differences, similarities, and time evolution. *International Journal of Educational Technology in Higher Education*, *12*(3), 98–112. https://doi.org/10.7238/rusc.v12i3.2515.

Lněnička, M., & Komárková, J. (2015). The impact of cloud computing and open (big) data on the enterprise architecture framework. In *Proceedings of the 26th IBIMA conference*, (pp. 1679–1683). Norristown: IBIMA.

Lněnička, M., & Komárková, J. (2019). Big and open linked data analytics ecosystem: Theoretical background and essential elements. *Government Information Quarterly*, *36*(1), 129–144. https://doi.org/10.1016/j.giq.2018.11.004.

Lněnička, M., & Máchová, R. (2015). Open (big) data and the importance of data catalogs and portals for the public sector. In *Proceedings in global virtual conference: The 3rd international global virtual conference (GV-CONF 2015)*, (pp. 143–148). Zilina: EDIS - Publishing Institution of the University of Zilina.

Lněnička, M., Máchová, R., Komárková, J., & Čermáková, I. (2018). Big and open linked educational data analytics: A research on stakeholders' capabilities, skills, and attitudes. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *ICERI2018 proceedings – 11th international conference of education, research and innovation*, (pp. 9549–9558). IATED Academy.

Lněnička, M., Máchová, R., Komárková, J., & Pásler, M. (2017). Government enterprise architecture for big and open linked data analytics in a smart city ecosystem. In V. Uskov, R. Howlett, & L. Jain (Eds.), *Smart Education and e-Learning 2017. SEEL 2017. Smart innovation, systems and technologies, 75*, (pp. 475–485). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59451-4_47.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, *9*(2), 17–28.

Máchová, R., Komárková, J., & Lněnička, M. (2016). Processing of big educational data in the cloud using apache Hadoop. In *Proceedings of the international conference on information society (i-society 2016)*, (pp. 46–49). London: Infonomics Society. https://doi.org/10.1109/i-Society.2016.7854170.

Marjanovic, D., Milovanovic, M., & Radenkovic, B. (2014). Hadoop infrastructure for education. In *XIV international symposium on new business models and sustainable competitiveness*, (pp. 365–370). Belgrade: University of Belgrade.

Mason, J., Khan, K., & Smith, S. (2016). Literate, numerate, discriminate – Realigning 21st century skills. In *Proceedings of the 24th international conference on computers in education*, (pp. 609–614). Asia-Pacific Society for Computers in education.

Mikalef, P., Giannakos, M. N., Pappas, I. O., & Krogstie, J. (2018). The human side of big data: Understanding the skills of the data scientist in education and industry. In *2018 IEEE global engineering education conference (EDUCON)*, (pp. 503–512). IEEE. https://doi.org/10.1109/EDUCON.2018.8363273.

Mikroyannidis, A., Domingue, J., Maleshkova, M., Norton, B., & Simperl, E. (2016). Teaching linked open data using open educational resources. In D. Mouromtsev, & M. d'Aquin (Eds.), *Open data for education. Lecture notes in computer science, 9500*, (pp. 135–152). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-30493-9_7.

Miller, S. (2014). Collaborative approaches needed to close the big data skills gap. *Journal of Organization Design*, *3*(1), 26–30. https://doi.org/10.7146/jod.9823.

Millette, C., & Hosein, P. (2016). A consumer focused open data platform. In *Proceedings of the 2016 3rd MEC international conference on big data and Smart City: ICBDSC*, (pp. 1–6). IEEE. https://doi.org/10.1109/ICBDSC.2016.7460350.

Navarrete, R., & Luján-Mora, S. (2015). Use of linked data to enhance open educational resources. In *Proceedings of the 14th international conference on information technology based higher education and training (ITHET 2015)*, (pp. 1–6). IEEE. https://doi.org/10.1109/ITHET.2015.7218017.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management*, *42*(1), 15–29. https://doi.org/10.1016/j.im.2003.11.002.

Pecori, R. (2018). A virtual learning architecture enhanced by fog computing and big data streams. *Future Internet*, *10*(1), 4. https://doi.org/10.3390/fi10010004.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532.

Rowley, T. J. (1997). Moving beyond dyadic ties: A network theory of stakeholder influences. *The Academy of Management Review*, *22*(4), 887–910. https://doi.org/10.2307/259248.

Sedkaoui, S. (2018). How data analytics is changing entrepreneurial opportunities? *International Journal of Innovation Science*, *10*(2), 274–294. https://doi.org/10.1108/IJIS-09-2017-0092.

Self, R. J. (2014). Governance strategies for the cloud, big data, and other technologies in education. In *2014 IEEE/ACM 7th international conference on utility and cloud computing*, (pp. 630–635). IEEE. https://doi.org/10.1109/UCC.2014.101.

Song, I. Y., & Zhu, Y. (2015). Big data and data science: What should we teach? *Expert Systems*, *33*(4), 364–373. https://doi.org/10.1111/exsy.12130.

Van den Broek, T., van Veenstra, A. F., & Folmer, E. (2011). Walking the extra byte: A lifecycle model for linked open data. In E. Folmer, M. Reuvers, & W. Quak (Eds.), *Linked open data – Pilot linked open data Nederland*, (pp. 95–111). Amersfort: Remwerk.

Van der Waal, S., Węcel, K., Ermilov, I., Janev, V., Milošević, U., & Wainwright, M. (2014). Lifting open data portals to the data web. In S. Auer, V. Bryl, & S. Tramp (Eds.), *Linked open data – Creating knowledge out of interlinked data*, (pp. 175–195). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09846-3_9.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Journal of Educational Technology & Society*, *15*(3), 133–148.

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, *122*, 119–135. https://doi.org/10.1016/j.compedu.2018.03.018.

Williamson, B. (2016). Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, *31*(2), 123–141. https://doi.org/10.1080/02680939.2015.1035758.

Yu, X., & Wu, S. (2015). Typical applications of big data in education. In *2015 international conference of educational innovation through technology*, (pp. 103–106). IEEE. https://doi.org/10.1109/EITT.2015.29.

Zadeh, A. H., Schiller, S., Duffy, K., & Williams, J. (2018). Big data and the commoditization of analytics: Engaging first-year business learners with analytics. *E-Journal of Business Education & Scholarship of Teaching*, *12*(1), 120–137.

Zheng, Q., He, H., Ma, T., Xue, N., Li, B., & Dong, B. (2014). Big log analysis for e-learning ecosystem. In *2014 IEEE 11th international conference on e-business engineering (ICEBE)*, (pp. 258–263). IEEE. https://doi.org/10.1109/ICEBE.2014.51.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R. S., & Sheikh Alibaks, R. (2012). Socio-technical impediments of open data. *Electronic Journal of e-Government*, *10*(2), 156–172.

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, *19*(1,2), 17–33. https://doi.org/10.3233/IP-140329.

Publisher's Note