A Review on How Big Data Analytics Can Influence Education

Victor Chang[1][a], Qianwen Xu[1] , Victor Mendez[2]

[1] *International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, China*

[2]*Universitat Autònoma de Barcelona, UAB, Spain*

Victor.Chang@xjtlu.edu.cn, Qianwen.Xu18@student.xjtlu.edu.cn, victor.mendez@uab.es

Keywords: Big Data, Big Data Analytics, Education, Ethics

Abstract: This paper gives an explanation about the definition of Big Data, Big Data Analytics and how they influence education in terms of benefits and ethical issues. First of all, this paper illustrates the applications of big data analytics in educational context and three benefits to education. Then, it explains ethical issues arisen during the usage of Big Data Analytics. It then discloses the ethical issues of Big Data Analytics employed in the education context from the aspects of individuals and the goal of education. Finally, this paper provides several recommendations to deal with current or potential ethical issues due to the applications of big data analytics in the area of education.

1 INTRODUCTION AND BIG DATA ANALYTICS

Nowadays, with the development of big data, a lot of industries have benefited a lot from the applications of big data analytics. While big data generates information from a variety of sources and processes them to create new information or knowledge and then make predictions or other usage, concerns such as ethical issues also become the hot topic. This paper explains the role of Big Data and Big Data Analytics in the area of education. It intends to provide a holistic perspective of education-related literature and discusses about the ethical issues within the context in detail.

1.1 What Is Big Data?

Big data is a broad context that refers to the massive amounts of digital information being captured and used to personalize content, predict consumer behaviors, and design interventions (Gregg, Wilson and Parrish, 2018). There are many definitions about big data. Mills, S., et al (2012) define big data as vast amount of data which is also of high velocity, complexity and variety and advanced technologies are required to collect, store, distribute, manage and analyze the information.

After conducting surveys on twelve definitions by Gartner, Microsoft, Oracle, Intel etc., Ward and Barker (2013) condense them and define the characteristics of big data as large volume and complex and it is stored and analyzed with a set of technical tools such as NoSQL, Map Reduce, and machine learning.

Nowadays, it is quite common to describe big data by using the Three V's, which are Volume, Variety, and Velocity, Volume refers to the magnitude of data. According to Gandomi and Haider (2014), definitions of big data volume could be different depending on factors, such as time. At present, the size of big data is measured in multiple terabytes and petabytes, but as the storage capacities keep growing, the data which can be classified as 'big', may not meet the entry criteria in the future as the measurement unit of big data may come up to Exabyte. Variety refers to diversity in the structure of dataset. Dataset can be structured (e.g. tabular data), semi-structured, and unstructured data (e.g. text, audio). Extensible Markup Language (XML), which is a language for flexible encoding documents, is a kind of semi-structured data. Velocity refers to the speed of data's generation, analysis and reaction (Gandomi and Haider, 2014). More and more areas, e.g. markets, mobile apps, require real-time information in recent years and will be in the future. High-frequency data

[a] https://orcid.org/0000-0002-8012-5852

enables industries and companies to analyze and create customer value in real time.

Two other properties which are Veracity and Value have been added to describe big data. They are related with the data lifecycle, data curation or smart data pre-processing (Garcia et al., 2016). Together with Volume, Variety, and Velocity, these five properties make it Five V's of big data. Veracity indicates the need to check and correct inaccuracy of data. In some sources of data, such as Facebook, data can be incorrect as the information from Facebook usually contain personal perspectives which are often not neutral and cannot be reliable. As for the last property of big data, Value, Rajaraman (2016) introduces it as no value without processing to obtain information using advanced techniques. What is important is that the five properties of big data are related to each other, one property changes may result in other properties changing.

1.2 What Is Big Data Analytics?

From the discussion above regarding to five properties of big data, Value, it can be seen that big data are unreliable and useless alone when they are collected from original source. They are only valuable when they are managed, analyzed, interpreted and the outcomes are used to decision making. Labrinidis and Jagadish (2012) divide the entire process of drawing out insights from big data into two sub-processes (Figure 1), which are data management and data analytics. One sub-process, data management, includes three steps. The first one is acquisition and recording, the second is extraction, cleaning and annotation, and the last one is integration, aggregation and representation. The other sub-process is data analytics, it consists of two steps, which are modelling and analysis, and interpretation. There are several different types of big data analytics. Gandomi and Haider (2014) mainly discuss four types of data analytics from the perspective of data's structure and source, which are Text analytics, Audio analytics, Video analytics and Social media analytics while Rajaraman (2016) divides data analytics into Descriptive Analytics, Predictive Analytic, Exploratory or Discovery Analytics, Prescriptive analytics by purpose. Nowadays, fields such as the weather, transportation, business have employed big data analytics to guide decision making and some of them have benefited from it, and, more recently, data analytics have been applied to higher education as well. (Boyd & Crawford, 2012; Clow, 2013; Wilson, Thompson, Watson, Drew, & Doyle, 2017).

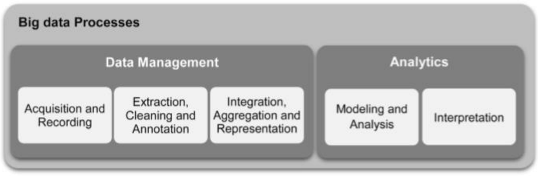

Figure 1: Processes for extracting insights from big data.

2 BIG DATA ANALYTICS IN EDUCATION

2.1 Learning Analytics

Big data applied to education is often referred to as learning analytics (LA). Buckingham Shum (2012) has divided learning analytics into three levels (see Figure 2):

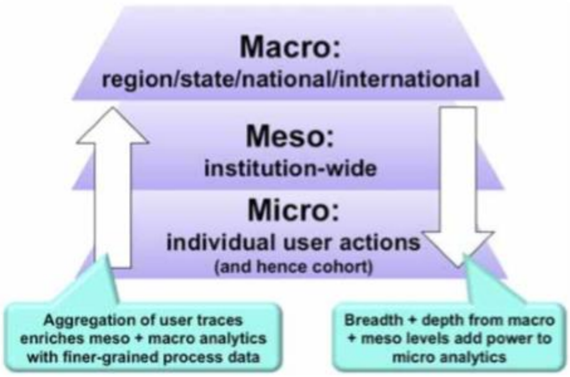

Figure 2: Three levels of learning analytics.

The first one is Macro level analytics. Macro level analytics make it possible for institutions to share data with each other for a variety of purposes, such as benchmarking. By using this analytics, education authorities and funding bodies are able to determine priorities for investment and development by identifying level patterns like geographical and socio-economic trends.

The second one is Meso level analytics. The target users of this analytics are usually individual institutions and the applications are mostly based on business intelligence approaches. It is quite common that problems of sharing data between centrally managed administrative systems and teaching and learning applications occur from time to time. For example, because of lacking the ability to exchange and use information, the information between student record systems and virtual learning environments cannot be shared and used effectively. Therefore, institutions should consider the institutional approaches at the strategic level, in terms of data sharing, re-use, management, and protection.

The last one is Micro level analytics. It is used to track data throughout individuals' whole learning processes and interpret the outcomes for individual learners to guide for their personal study. Powell and MacNeill (2012) have listed a variety of motivators for the application of learning analytics. Institutions use learning analytics most, they use them to share study materials, record attendance, turn in assignment, and they employ learning analytics to identify the students at-risk, so teachers and support staff are able to intervene earlier to provide extra support to help these students get back on rail. Learning analytics also enables functional groups, such as module teams, to improve current modules or develop new curriculums. For institutional administrators, learning analytics are useful as they provide information for them to make decisions on businesses, including marketing and recruitment. Analytics provide newer forms and services for business intelligence (Halper, 2014), since activities, performance, finance and customer feedback of the businesses can be recorded and presented by analytics. In terms of individual learners for education, learning analytics guide their personal study by reflecting on their achievements and comparing their patterns of performance to their peers.

Long and Siemens (2011) share a similar opinion with Buckingham Shum on data analytics in education context. However, they divide the learning analytics into two categories, academic analytics and learner analytics. The academic analytics correspond to macro and meso level analytics while learner analytics to micro level analytics, similar to viewpoints presented by Halper (2014). In Long and Siemens' opinion, the concentration of learning analytics is on the learning process in specific.

2.2 Educational Content Analytics

Apart from being used in teaching and learning processes, data analytics are also useful in educational content and institutions apply this kind of analytics in resource management systems. The basis of educational content analytics is paradata. Campbell & Barker (2013) defines paradata as data on what kind of educational resource is used, who is using and how they are used. In order to manage and share information on learning resources more effectively, it is quite important for educational content sharing system to identify the features that are possibly attractive or relevant to their users. The systems used to make use of educational metadata to identify all these features but it struggles to do so because learning resources represent a diverse class of objects, used in a wide variety of contexts but metadata can only collect data related to properties of the learning resource itself instead of all the features the systems require (Barker & Campbell, 2010). But paradata is different, as it can not only to classify the properties of the learning resource itself, and instead involves the capture of in situ information about online users' actions related to the resource, such as viewing, downloading, sharing, favoring, and embedding reusable content into derivative works, adding tags, and so on. Therefore, educational content analytic, in the form of paradata, is considered to be valuable in the term of generation and analysis about data on how educational resources is used in real life.

3 HOW BIG DATA ANALYTICS BE BENEFICIAL TO HIGHER EDUCATION?

Staff and students can be benefited positively by data analytics in education in a number of ways. By making use of data analytics in education context, in addition to the convenience and efficiency, institutions are able to improve students' performance, enhance their study experience, but also increase graduation rates and improve teaching methods.

3.1 Convenience

Nowadays, some institutions have employed learning management systems (LMS), such as Blackboard, ICE, Moodle to provide service for the convenience of faculty and students. On one hand, Professors can use LMS to upload lecture materials, assessments for students, record students' attendance, mark students' assessment, and release announcements to inform students with the latest information. On the other hand, students can easily obtain study materials to preview or review online and turn in their assignment just at home. Using learning management systems improves the students experience during their learning processes and improve the efficiency of faculty's work. Additionally, Business Intelligence can access business databases decoupled from the Learning Management Systems (LMS), so that users can retrieve information easily by Analytics tools.

The can enable 5V, convenience and easy accessibility to the Universities.

3.2 Improvement of Student Success

Big data analytics in higher education are not only convenience, but also useful to improve students' performance. Georgia State (Dimeo, 2017) uses LMS to collect dataset of learning activity and academic performance to identify students at-risk at the individual level while Sinclair Community College employs a Web-based counseling system to make an Individual Learning Plan (ILP) for students satisfying certain characteristics since 2004 (Campbell, DeBlois & Oblinger, 2017). Since 2012, Georgia State established an adviser system which employs the tool of learner analytics that keeps recording students' grades and information related to academic performance continuously and then analyze each of them with about 800 factors to identify the students who have high possibility to fail in graduation. Then, encouraging emails or system messages will be sent by the advice system with advice or helpful materials to support the students at-risk (Dimeo, 2017).

In the meantime, Sinclair Community College employed a Web-based counseling system called Student Success Plan (SSP) to initiate early alerts and enable advisers to make an ILP for students identified at-risk by the SSP (Campbell, DeBlois & Oblinger, 2017). The SSP is based on SQL database and enable to manage, refer records and to report. Students will be assessed in four aspects: grades, financial status, work and undecided major. Once an at-risk student is identified and a learning plan is created, additional information will be captured, such as course enrollment plan, tutorial materials, counselor's notes and so on. After employing the SSP, it shows that the rate of retention from freshman to sophomore year was 93.3 percent for students who completed ILP, 76 percent for those who activated but not completed ILP and 65.7 percent for those who did not activated ILP. The result indicated that the SSP contributes to the improvement in students' performance and retention.

3.3 Improvement of Teaching Methods

Teacher supporting tools used in the digital learning environment are the application of learning analytics as well. These tools can collect and analyze educational datasets and then report to students and teachers in order to optimize learning. By making use of supporting tools, it is more convenient for teachers to obtain the feedback from the educational activities. As a result, the information load may be lower and then teachers can concentrate on helping students rather than spending their time on analyzing the educational outcomes by themselves. In addition, up-to-date data from supporting tools enables teachers to conduct an intervention or adjust their teaching methods in time (Van Leeuwen et al. 2014).

4 ETHICAL ISSUES

Nowadays, big data analytics has been wide employed in a number of industries, including healthcare, electricity, law enforcement and many other fields. As the activity of generating, using and processing of data is getting more and more common, relative legal and ethical concerns have also caused the public's attention. Mason (1986) categories ethical issues on data into four aspects:

- Privacy –What information is held about the individual? Are there any safeguards?
- Accuracy –Is it accurate?
- Property –Who really owns this information?
- Accessibility –Who has the permission to access this information, and under which circumstances?

Privacy is the primary factor that concerns the public most, as nowadays, the problem of personal information leakage is so severe that almost everyone receives different kinds of phone scam or e-mail scam every day. Misinformation can cause massive damage to people's reputation, property and even lives (Mason, 1986). As for property, because information is flow and communicable, it is difficult for the individual to keep the information to oneself, so the ownership of the information is debatable. With development of data techniques, there are huge amount of and different kinds of information available online. In order to use these large data set to its maximum effect, some companies make profits from selling access to information, even some of the information is personal and without consent (Martin, 2015). Although this kind of activity is forbidden by the Data Protection Acts (Act, 1998).

Davis and Patterson also discussed their opinions on ethical issues resulted from big data analytics in the interview conducted by Howard Wen (2012), 'Big data itself, like all technology, is ethically neutral. The use of big data, however, is not.' Aside from the four categories Mason summarized, they considered liability (Who is ultimately responsible for maintaining it) is important as well.

Laws constrain the people and the society from doing things not acknowledged by them with enforcement. By contrast, ethics tells people what is supposed to do and not supposed to do in rather than must not do defined by laws (Chang and Lin, 2017). Although commercial vendors are introducing more and more analytics tools into different industries, these applications are still in infancy and may contain a variety of ethical challenges (MacNeill, Campbell and Hawksey, 2014) Therefore, further exploring and identifying is essential for stakeholders related to these collected data. In the rest of this paper, it will discuss the particular ethical issues in the context of higher education.

5 ETHICAL ISSUES IN HIGHER EDUCATION

5.1 Individuals

5.1.1 Privacy

Since its initial step of development, a lot of concerns have been raised on data analytics. The most important and common one is privacy. In the progress of data analytics, if the opportunity for consent to data activities, such as collection or use, is absent, a subject's privacy will be violated and his/her private information may be used in ways against their initial intentions. (Nissenbaum, 2014). When our privacy is violated, we may be uncomfortable about our information being divulged and get harassed by crank calls, or worse our information is manipulated of or discrimination against us.

Furthermore, the possibility of privacy violation in learning analytics and educational content analytics is at a high level too. Some institution collects and analyzes students' information for its own interest instead of students'. For example, Arizona State University (Parry, 2012) analyzes students' information to identify student who has intention to change university, so it can take action a step ahead to convince them to stay, but one ASU student reported (Johnson,2014) that information like this is not the kind of those students would want to share with universities.

Specific individuals or classes of individuals are also harmed though discrimination by some personalized learning techniques in education institutions. Discrimination can be divided into a lot of categories: race, religion, color, sex, age, marital status and so on. In the case of University of Alabama, in order to improve student retention, the university gave students who selected data-mining course access to the information of freshmen and asked them to develop a predictive model to identify students at-risk. The model considered eight variables, including the race of enrolled freshmen, which may lead to discrimination if the information is misused (Campbell, DeBlois & Oblinger, 2017). When specific individuals or classes of individuals' information is divulged, secret factors may be embedded in the algorithmic of the applications and make them arbitrary and unaccountable, and then lead the personalized learning techniques to put these individuals or classes of individuals in the incorrect group falsely or in a protected group, or deprive them of equal opportunities, or burden them with additional stress (Hildebrandt and Koops, 2010).

In addition, some predictive analytics used in institutions to help schools identify at-risk students who is about to fail in courses or even graduate from schools, so schools are enable to step in in time. In Hamilton County Board of Education in Tennessee, for example, the average scores of math and reading have went up by over 10 percent and the rate of graduation has arouse by over 8 percent since the introduction of IBM's Predictive Analytics Solution for Schools and Educational Systems (PASSES) (IBM, 2011). This technique may improve students' performance, but it also brings some ethical risks. The outcomes may be incorrect and the errors are difficult to recognize and reverse, so the educational opportunities may be restricted incorrectly. Meanwhile, students who are identified as having difficulties to graduate might be marked as less capable, and they might be considered or treated as if they are less capable by teachers or even by themselves (MacCarthy, 2014).

5.1.2 Individually

Challenges to individually can be raised by the applications of data analytics as well. Older techniques treat subjects all typical and try to analyze central tendencies or relationships among them as a whole while data mining makes individuality much easier as it identifies different characteristics among them. In education institution, learning analytics tend to treat students a collection of characteristics or attributes rather than a whole subject with their own opinions and step in to treat students as a predictive value or category. Haimson (2014) argues against the data collection activity initiated by inBloom because he thinks that inBloom's subjecting children to software programs makes education depersonalized, which is not moral and against the value of education. Applications of this kind of predictive analytics such as course recommendation systems do not consider students' own interests, they just match students' characteristics or skills to previous students' information and recommend them to choose the courses students similar to them have chosen before. Arizona State's eAdvising system (Parry, 2012), for example, identifies students whose "ambitiousness is not related to their abilities". This system is an obvious one that regard students only a set of skills rather than individuals who have their own feelings, thoughts, interests or ambitions.

5.1.3 Autonomy

Under extreme circumstances, the problem, individuality, in learning analytics could pose infringement on subjects' autonomy. Autonomy is one's ability or freedom to make or modify his or her own decisions without coercion and paternalism of others'. People with autonomy is able to act in accordance with their own judgements and preferences rather than being controlled or influenced by anyone else (Wikipedia, 2018). The reason why the problem of individuality might cause the challenge in autonomy of students is because learning analytics treat a subject just bulks of skills instead of a person, so its outcomes or suggested choices are what the techniques consider the wise or efficient for data users. In education institution, infringement on students' autonomy is at a very high level. Some course recommendation systems not just mean to provide suggested advice, but also to remedy students' poor decision-making (Parry, 2012). They lead students to choose the courses the institution wants them to choose, which are easier to pass by offering students multiple choices but quite little information about the courses (sometimes just some course descriptions). When students sign up for courses, they have no idea about the implications of taking different courses with limited information, so they make decisions based on some inessential factors or by considering which one is easier than others. In addition, some adaptive learning software violate students' autonomy more directly. They just tell students what to do in every step instead of responding to their questions. To be fair, violating autonomy is not intrinsically unacceptable, as, sometimes, it guides the students who are not mature enough and might make quite ill-considered decisions by themselves. But it is absolutely necessary to maintain the autonomy in education institution as one of the roles of schools is to teach competencies rather than mark on students' performance in courses (Johnson, 2014).

5.2 Harm to the Goal of Education

The interests of different stakeholders vary with each other on personal information. Students and parents want their information to be kept and used within the range they have permitted. Therefore, they will not be willing to share unless they make sure that their information is not going to be used incompatible with their intention. However, some employers, advertisers, and insurance companies might want their information for business decisions to make profits and this will undermine students and parents' trust, and then they will respond by avoidance, hostility and withdrawal. In this way, the possibility of information's misuse may decrease, but without sufficient amount of academic information from students, institutions cannot be able to improve their services in terms of course design, assessment, personalized learning program and so on. And Education Bureau will not able to set standard or strategy for institutions as well. Therefore, keeping students' information within permitted range is crucial and institution should only make student information accessible to third parties with consent from students or their parents (MacCarthy, 2014).

6 CASE STUDY: COURSE SIGNALS AT PURDUE UNIVERSITY

Purdue University (Arnold & Pistilli, 2012) employs a student success system called Course Signals to alert faculty and students to behavior that may lead to failure at the course level. The Course Signals uses learning analytics to predict students' performance in a given course and alert them with green (great), yellow (potential problem) or red (high risk of failure) light on their course homepage. This system identifies students who might be struggling through four factors: the points earned to date; time on task, compared to their classmates; past performance; and students' characteristics. If a student's light is yellow or red, the instructor then sends a personal e-mail or organize a face-to-face meeting to guide the student to support resources and tell him/her how to use them. Receiving feedback on their performance on a given course at regular intervals gives students the opportunity to keep themselves on track or adapt their behaviors as soon as possible.

After employing the Course Signals, students and faculty both benefited from the application of learning analytics. On one hand, the rate of success in individual courses went up. The number of As and Bs awarded increased by about 10 percent and the number of Cs awarded decreased by nearly 6.5 percent among the users of Course Signals with a comparison to previous semesters which provided the same courses but didn't used Course Signals. In addition, the increasing success rate in individual courses resulted in an increasement in students' retention to graduation as well. On the other hand, from the perspective of faculty, this system allowed teachers to adapt their teaching methods in time as well. For example, if they find most of their students have red lights, then they are able to communicate with them to find out the problem with their course design or teaching methods ahead of the end of the semester (Arnold & Pistilli, 2012).

Although the system of Course Signals brings a number of advantages to students and faculty, there are some potential ethical issues that should be considered. Firstly, students may be threatened by privacy issue. Institutions have the obligation to inform students that their footprints or behaviors in the applications of learning analytics can be traced by Course Signals and only collect their information when receive their formal consents. Meantime, students should have the opportunity to opt out. Secondly, the information Course Signals collect may lack of accuracy. For example, in the real world, a great amount of learning activities take place outside the institutions but Course Signals can only be able to measure students' efforts within the institution in term of students' interaction with Blackboard Vista, Purdue's learning management system. Therefore, the information the Course Signals collect may be incomplete and cannot represent a holistic view of a student and the prediction based on this kind of information may be misleading and inaccurate. Finally, the interpretation of data may vary from analysts to analysts as different analysts have different perspectives and the predicted results may be biased (Slade & Prinsloo, 2013).

7 RECOMMENDATIONS

Adoption of educational analytics with ethical concerns can be effective yet challenging. Therefore, this section describes a list of recommendations. First, Kay, Korn and Oppenheim (2012) provide several principles for good practices:

- Clarity: The policy should be comprehensive and clarified;
- Comfort and care: Institutions should consider the benefits of data-providers during the process of collecting and reusing, including unusual cases;
- Choice and consent: Institutions should provide opt-in or opt-out choices and the opportunity to withdraw at any time for data-providers;
- Consequence and complaint: In order to protect students' privacy, Institutions should de-identify students' data and prepare for unpredictable influences and establish a redress system for data-providers.

Foremost, this paper considers the employment of any learning analytics in education should align with the value or purpose of education. Institutions should always keep its mission in mind as teaching competencies and cultivating talents for business world instead of marking students on their performance and compete with other institutions by high graduation rate (MacCarthy,2014).

Second, the clarified and comprehensive ethical principles or policies should be established in terms of aims, range and boundaries. And these principles or policies should provide to students or parents before they choose to opt in or opt out, making sure they are fully aware of what they are doing and the rights they have.

Third, institutions should not only provide opt-in or opt-out options for students or parents, but also provide them the opportunity to withdraw at any time when they feel their data is infringed or any other reasons.

Fourth, during the process of generating, reusing data of students, institutions should take students' interest and feeling into account and prepare for unpredictable impacts arising from holding so much data. Therefore, it is better for institutions to establish a usable and reliable system for students to complain and require for revoking their information.

Last but not least, we can make an interactive, and learning community blending the offline and online learning. A teaching and ethics committee will be established to lead and steer all the recommended areas above, together with the students, and faculty members, administrators and policy makers. By following these recommendations, it is much easier and convenient for the institutions to check and update students' records of academic performance.

The careful handling of data can also meet the Data Protection Act and also the careful use of data meeting the legal requirements.

8 CONCLUSION AND FUTURE WORK

There is no doubt that data analytics is conducive to higher education and it is an evitable trend that more and more applications will be used in education because students who have difficulties in learning require guidance customized according to their own circumstances from these applications and from the aspect of resource utilization, it is not efficient to allow a great deal of students to fail in graduation. Therefore, the function of big data analytics to improve the efficiency of higher education cannot be denied.

However, the current of potential ethical issues or challenges cannot be ignored as well. From this paper’s discussion above, the applications of big data analytics in education cause several ethical issues, including privacy, individuality and autonomy. In order to reduce these issues, firstly, it is essential for institutions to align the aims and goals of using these applications with the value of education. Secondly, institutions can offer students or parents opportunities to opt-in, opt-out or withdraw at any time and provide them a redress system.

Our future may include developing a framework that can integrate the best practices in all the aspects of learning and the use of learning analytics. We aim to improve the efficiency, quality and convenience of learning, teaching methods, students’ learning experience, as well as a better handling in new and existing ethical concerns.

REFERENCES

Arnold, K. E., & Pistilli, M. (2012). Course signals at Purdue: Using learning analytics to increase student success. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 267–270.

Act, D. P. (1998). Hmso.

Autonomy. (2018) Retrieved from https://en.wikipedia.org/wiki/Autonomy

Barker, P., & Campbell, L.M. (2010). Metadata for learning materials: An overview of existing standards and current developments. Technology, Instruction, Cognition, and Learning, 7(3-4), 225-243.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662-679.

Buckingham Shum, S. (2012). Learning Analytics. UNESCO Policy Brief. Retrieved from http://iite.unesco.org/pics/publications/en/files/3214711.pdf

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. EDUCAUSE review, 42(4), 40.

Campbell, L. M., & Barker, P. (2013). Activity Data and Paradata. Cetis Briefing Paper, 2013: B01. Retrieved from http://publications.cetis.ac.uk/2013/808

Chang, V. I., & Lin, W. (2019). How Big Data Transforms Manufacturing Industry: A Review Paper. International Journal of Strategic Engineering (IJoSE), 2(1), 39-51.

Clow, D. (2013). An overview of learning analytics. Teaching in Higher Education, 18(6), 683-695.

Dimeo, J. (2017). Georgia State improves student outcomes with data. Inside Higher Ed. Retrieved from https://www.insidehighered.com/print/digitallearning/article/2017/07/19/georgia-state-improves-studentoutcomes-data

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. Big Data Analytics, 1(1), 9.

Gregg, A., Wilson, B. G., & Parrish, P. (2018). Do No Harm: A Balanced Approach to Vendor Relationships, Learning Analytics, and Higher Education. IDEA Center, 72.

Haimson, L. (2014). Statement on inBloom's overdue demise. Retrieved from http://nycpublicschoolparents.blogspot.com/2014/04/statement-on-inblooms-demise.html

Halper, F. (2014). Predictive analytics for business advantage. TDWI Research, 1-32.

Hildebrandt, M., & Koops, B. J. (2010). The challenges of ambient law and legal protection in the profiling era. The Modern Law Review, 73(3), 428-460.

IBM. (2011) Mobile County Public Schools: Analytical insights help keep students on track. Retrieved from http://www.ibm.com/smarterplanet/us/en/leadership/mobilecounty/assets/pdf/IBM_MobileCounty.pdf

Johnson, J. A. (2014). The ethics of big data in higher education. International Review of Information Ethics, 21(21), 3-10.

Kay, D., Korn, N., & Oppenheim, C. (2012). Legal, risk and ethical aspects of analytics in Higher Education. Cetis Analytics Series, 1(6).

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5). Retrieved from http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education.

MacCarthy, M. (2014). Student privacy: Harm and context.

MacNeill, S., Campbell, L. M., & Hawksey, M. (2014). Analytics for education. Reusing Open Resources: Learning in Open Networks for Work, Life and Education, 154.

Martin, K. E. (2015). Ethical issues in the big data industry.

Mason, R. O. (1986). Four ethical issues of the information age. Mis Quarterly, 5-12.

Mills, S., Lucas, S., Irakliotis, L., Rappa, M., Carlson, T., & Perlowitz, B. (2012). Demystifying big data: a practical guide to transforming the business of government. TechAmerica Foundation, Washington.

Parry, M. (2012). College degrees, designed by the numbers. The Chronicle of Higher Education.

Powell, S., & MacNeill, S. (2012). Institutional readiness for analytics. Cetis Analytics Series, 1(8). Retrieved from http://publications.cetis.ac.uk/2012/527

Rajaraman, V. (2016). Big data analytics. Resonance, 21(8), 695-716.

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

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. Computers & Education, 79, 28-39.

Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions.

Wen.H. (2012). Big ethics for big data. Retrieved from https://www.oreilly.com/ideas/ethics-big-data-business-decisions

Wilson, A., Thompson, T. L., Watson, C., Drew, V., & Doyle, S. (2017). Big data and learning analytics: Singular or plural?. First Monday, 22(4).