British Journal of Educational Technology *Vol 46 No 5 2015* 904–920
doi:10.1111/bjet.12230

Big Data and analytics in higher education: Opportunities and challenges

Ben Daniel

Dr. Ben Daniel is a Senior Lecturer in Higher Education, and heads an Educational Technology Group, at the University of Otago—New Zealand. His current research is focused on understanding the value of Big Data and learning analytics in higher education. He is also investigating theories and praxis of teaching research methodologies for Business and Academia.

Abstract
Institutions of higher education are operating in an increasingly complex and competitive environment. This paper identifies contemporary challenges facing institutions of higher education worldwide and explores the potential of Big Data in addressing these challenges. The paper then outlines a number of opportunities and challenges associated with the implementation of Big Data in the context of higher education. The paper concludes by outlining future directions relating to the development and implementation of an institutional project on Big Data.

Introduction

Institutions of higher education are operating in an increasingly complex and competitive environment. They are under increasing pressure to respond to national and global economic, political and social change such as the growing need to increase the proportion of students in certain disciplines, embedding workplace graduate attributes and ensuring that the quality of learning programmes are both nationally and globally relevant.

In addition, different stakeholders are expecting higher education institutions to in a timely manner to these demands, albeit with declining government funding, declining support from business and private sectors, growing regulatory demands for transparency and accountability (Hazelkorn, 2007), declining admissions rates due to increasing tuition and upsurge in high schools dropout and increasing operational costs (Thornton, 2013).

How can institutions of higher education respond effectively and in time to global changes affecting their environment? The decisions required for dealing with these rapid changes are complex and many are made without recourse to vast data sources that have been generated but are not available to those entrusted to make relevant and timely choices. These data can play a major part in how we understand the often contested nature of higher education governance (Clarke, Nelson & Stoodley, 2013) and so ensure that institutions are not only able to respond effectively to changes happening within and outside them, but that they also remain pertinent to their purpose in the societies that they serve.

This paper examines the role of Big Data and Analytics in addressing contemporary challenges facing higher education (see Figure 1). The paper first identifies the key global trends affecting institutions of higher education and explores the potential of Big Data and Analytics in addressing these changing trends. Secondly, the paper outlines opportunities and challenges associated with the implementation and governance of Big Data in higher education. The paper concludes by outlining future directions relating to the development and implementation of institutional project on Big Data.

Practitioner Notes

What is already known about this topic

- Big Data is an emergent field of research that uses data analysis to inform decisions. It is currently being explored mostly in business, government and health care due to the growing plethora of data collected and stored in these environments.
- Big Data research is mainly aimed at examining how to efficiently aggregate and correlate massive volumes of data to identify recurring behavioural patterns and meaningful trends instead of cataloguing the status quo.

What this paper adds

- There is limited research into big data in higher education, despite growing interests in exploring and unlocking the value of the increasing data within higher education environment.
- This paper contributes to the conceptual and theoretical understanding of Big Data and Analytics within higher education.
- It introduces the notion of Big Data and outlines its relevance to higher education.
- It describes the opportunities this growing research area brings to higher education as well as major challenges associated with its exploration and implementation.

Implications for practice and/or policy

- It outlines major challenges faced by institutions of higher education and emphasises the role of Big Data in addressing some of these challenges.
- The paper also describe how Big Data can be utilised to inform policy, programme and teaching, learning and research outcomes.
- Opportunities and challenges discussed in the paper will enable policymakers and information technology analyst to make informed choices when considering to explore and implement Big Data programmes in their institutions.

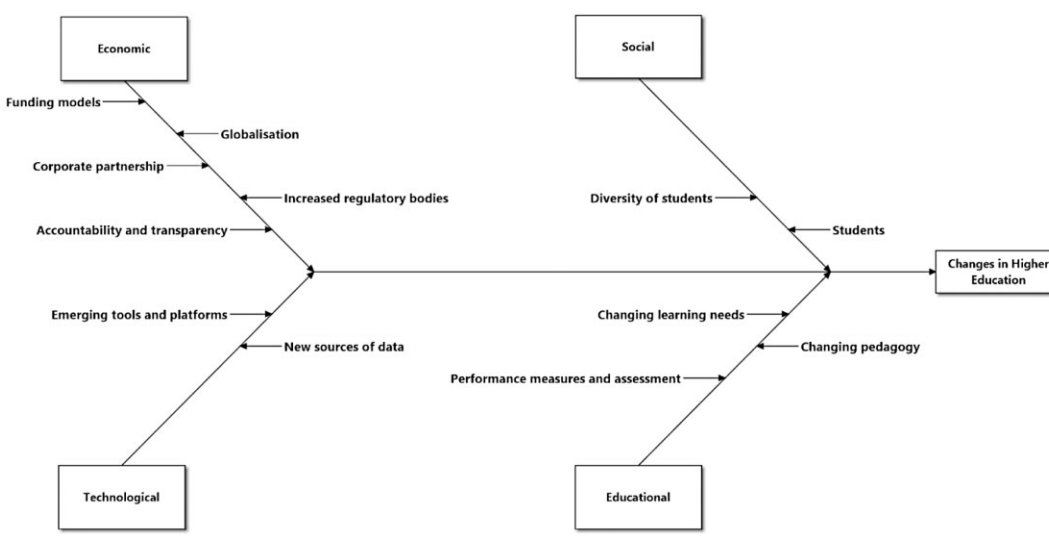

Figure 1: Current trends affecting institutions of higher education

The paper also opens up new research areas that can be explored to enrich our understanding of the role of Big Data in higher education.

Corporate–academic partnerships are increasing (Leydesdorff & Etzkowitz, 2001); however, to attract and sustain these partnerships, corporations require institutions of higher education to demonstrate a commitment to the utilisation and development of advanced technologies that are likely to support applied research outputs and potentials for knowledge transfer and commercialisation (Mok, 2005).

At the same time, these global changes are mounting on the institutions of higher education; new technologies continue to have a significant impact on academic careers as research and teaching become more reliant on these technologies (Economist Report, 2008). Over the last decades, a digital revolution associated with developments in new technologies such as ubiquitous computing devices, flexible class room design and Massive Open Online Courses is radically reshaping the mode and accessibility of learning and teaching. In addition, many institutions are embracing new class formats and technologies designed to meet either evolving student needs or as mechanisms to reduce operational costs.

In spite of the growing changes happening in the environment of higher education, the role of data in helping addressing contemporary challenges is often overlooked. As learning technologies continue to penetrate all facets of higher education, a plethora of useful 'data traces' are generated. These data can be utilised to inform institutions of higher education to adapt better in response to changes happening within and outside their environments.

Knowledge discovery and data mining approaches

Knowledge discovery (KDD) is an interdisciplinary area focusing on methodologies for identifying and extracting useful and meaningful patterns from large data sets. KDD draws upon research in statistics, databases, pattern recognition, machine learning, data visualisation, optimisation and high performance computing.

First proposed in early 1990s, data mining, ie, applying data analysis and discovery algorithms to produce a particular enumeration of patterns (or models) over the data, became a major aspect of KDD. Over the years, clustering, association, classification algorithms, regression models, predictive methods and factor analysis are the key approaches that have come to dominate data mining research.

Clustering algorithms or data segmentation approaches group data items based on clearly defined logical relationships. The goal is to maximise the inter-cluster similarity and minimise the intra-cluster similarity. Related to data clustering is feature selection. Feature selection refers to the process of identifying the most useful feature (or combinations of features) for clustering.

Classification, is another approach within the data mining research that tends to classify data into predetermined categories. Classification approaches are often referred to as machine learning algorithms. These algorithms can learn from a large set of pre-classified data. They can detect persistent systemic differences between items in each group and apply particular rules to further classify data.

Association approaches are also a set of algorithms intended to extract particular characteristics of data within a group and to finding associations with other characteristics. These approaches are driven by rule-based algorithms, which mostly examine relationships within data set (correlation between variables). The goal of association rules mining is to study the frequency associated with relationships among items or data in transactions (Agrawal, Imielinski & Swami, 1993). A simple association rule expresses the relationship of one item to another (ie, simple linear relationship).

Regression is perhaps one of the most commonly used data mining approaches for constructing predictive models. Linear regression is similar to linear models used in statistical significance testing.

Baker (2013) indicated that predictive models in educational data mining context are intended to infer a single aspect of the data (the predicted variable, similar to dependent variables in traditional statistical analysis) from some combination of other aspects of the data (predictor variables, similar to independent variables in traditional statistical analysis).

Factor analysis is also a common approach in education, its goal is to find variables that can be naturally grouped together, or splitting the set of variables (as opposed to the data points) into a set of latent (not directly observable) factors (Kline, 1993). Within educational settings, factor analysis includes algorithms such as principal component analysis and exponential-family principal components analysis and used for dimensionality reduction (eg, reducing the number of variables), including in pre-processing to reduce the potential for over-fitting and to determine meta-features. Factor analysis has also been used in learning environments to develop predictive models (Baker & Yacef, 2009; Minaei-Bidgoli, Kashy, Kortmeyer & Punch, 2003).

The data mining approaches described above are prominent in the work of Romero and Ventura (2010), who applied these in educational data mining research. These approaches draw on a variety of literatures, including data mining and machine learning, psychometrics and other areas of statistics, information visualisation, and computational modelling (Baker & Yacef, 2009). Unlike traditional data mining approaches most of which are aimed at automating the process for detecting interpretable patterns and carrying out predictions, educational data mining tends to focus on developing new tools for discovering patterns in data and applying tools and techniques to analyse large sets of data (Luan, 2002; Romero & Ventura, 2010).

Emergence of Big Data

Many organisations are currently using data to make better decisions about their strategic and operational directions. Using data to make decisions is not new; business organisations have been storing and analysing large volumes of data since the advent of data warehouse systems in the early 1990s. However, the nature of data available to most organisations is changing, and the changes bring with them complexity in managing the volumes and analysis of these data. Basu (2013) observed that most businesses today run on structured data (numbers and categories). However, this does not reflect the complexity on the nature of available corporate data and their untapped hidden business value. According to IBM, 80% of the data organisations currently generated are unstructured, and they come in a variety of formats such as text, video, audio, diagrams, images and combinations of any two or more formats.

Most of these unstructured data make their way to corporate data warehouse. The term 'Data warehouse' refers to a central repository of data or a centralised database. It represents an ideal vision of maintaining a central data repository and a living memory of data that can be leveraged for better decision making.

Recent developments in database technologies made it possible to collect and maintain large and complex amounts of data in many forms and from multiple sources. In addition, there are analytical tools available that can turn this complex data into meaningful patterns and value, a phenomenon referred to as Big Data.

Big Data describes data that is fundamentally too big and moves too fast, thus exceeding the processing capacity of conventional database systems (Manyika *et al*, 2011). It also covers innovative techniques and technologies to capture, store, distribute, manage and analyse larger sized data sets with diverse structures.

With new concepts, critiques emerge. Some critics contested that the notion of Big in the term itself is misleading and that it does not reflect only data size but complexity. Yang (2013) pointed out the definition of Big Data has little to do with the data itself, as the analysis of large quantities of data is not new, but rather Big Data refers to emergent suit of technologies that can process mass volumes of data of various types at faster speeds than ever before. This conceptualisation of Big Data was echoed by Forrester defining Big Data as "technologies and techniques that make capturing value from data at an extreme scale economical." The term economical suggests that if the costs of extracting, processing and making use of data exceed the advantages to be collected, then it is not worth indulging in the process. Generally, Big Data has come to be identified by a number of fundamental characteristics. Key among them are:

- Volume—large amount of information is often challenging to store, process, and transfer, analyse and present.
- Velocity—relating to increasing rate at which information flows within an organisation—(eg, organisations dealing with financial information have ability to deal with this).
- Veracity refers to the biases, noise and abnormality in data. It also looks at how data that is being stored, and meaningfully mined to the problem being analysed. Veracity also covers questions of trust and uncertainty.
- Variety—referring to data in diverse format both structured and unstructured.
- Verification—refers to data verification and security.
- Value—most importantly, has the data been utilised to generate value of the insights, benefits and business processes, etc. within an organisation?

Douglas (2001) in the Gartner's report proposed the three of the most common properties of Big Data. The report made three fundamental observations: the increasing size of data, the growing rate at which it is produced and the cumulative range of formats and representations employed, proposed threefold definition encompassing the "three Vs" (Volume, Velocity and Variety).

There are also other properties of Big Data such as data validity, which refers to accuracy of data, and volatility, a concept associated with the longevity of data and their relevance to analysis outcomes, as well as the length required to store data in a useful form for appropriate value-added analysis. In addition to these properties, there are three stages required to unlock the value of Big Data in any organisation. These include data collection, data analysis, visualisation and application (see Figure 2).

Collection

Data collection is the first step in unlocking the value accrued from Big Data. This requires identifying data that can reveal useful and valuable information. Data must be filtered for relevance and stored in a form that is useful, as little is gained in investing in huge amounts of data and storage infrastructure if the vast majority of the data in it is not usable.

Analysis

Once data have been rendered into a usable form, it has to be analysed to generate actionable information. However, with the growing diversity in the nature of data, managing and analysing diverse data set is becoming a very complex process. Analysis needs to include linking, connecting correlating different data sets to be able to grasp the information that is supposed to be conveyed by these data. This situation is, therefore, termed as the 'complexity' of Big Data.

Visualisation and application

This is the last stage where the analysed data is made available to users in a form that is interpretable and integrated into existing processes, and ultimately used to guide decision making.

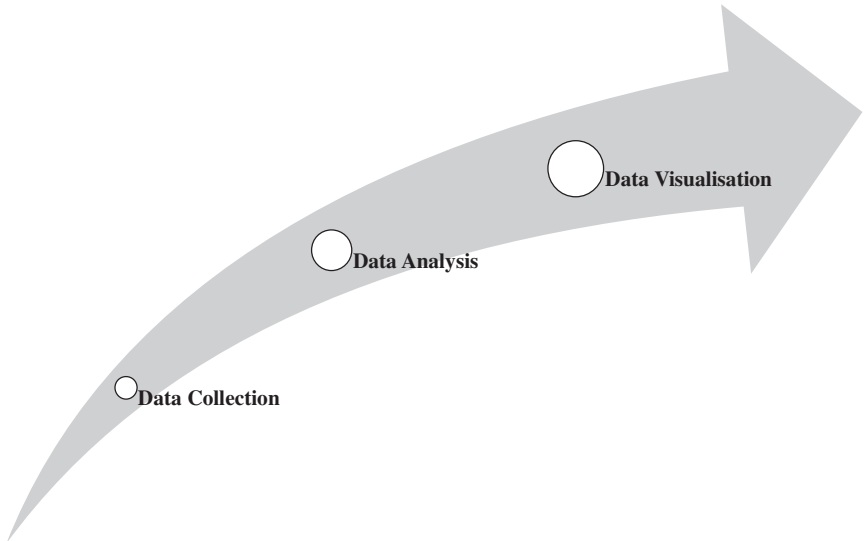

Figure 2: Three essential stages of Big Data

Analytics and data warehousing

Trillions of data are being collected and stored in various institutional databases. Many institutions of higher education are increasingly, delivering learning online. Subsequently, there is a widespread availability of online repositories, educational digital libraries, and their associated tools, all of which could be utilised as major catalysts for change in practice (Borgman *et al*, 2008; Choudhury, Hobbs & Lorie, 2002; Xu & Recker, 2012). Data is stored in student information systems, student social media usage data, learning management systems, student library usage, individual computers and administrative systems holding information on programme completion rates and learning pathways.

In addition to growing abundance in data, the data come in different formats of audio, video, text and pictures. Furthermore, when learners interact with a digital device, data about that interaction can be easily captured or "logged" and made available for subsequent analysis.

The variety of data collected and stored could explored using modern analytic techniques. Analytics are attractive approaches in education due to ability of tools and techniques for data processing and analysis. Mayer (2009) noted that the increase in attention to analytics is also driven by advances in computation, especially on various platforms. For instance, smartphones currently exceed the computational power of desktop computers and are more powerful than mainframe computers today which can accomplish tasks that were impossible only a few years ago (Baker & Inventado, in press). In addition, increases in computational power support researchers in analysing large quantities of data. In addition, the appearance of new systems and analytical tools are the most significant advances in research in analytics.

There is also a growing number of systems and tools that are intended to leverage Big Data and analytics. For instance, systems such as Apache Hadoop, Hortonworks, MapReduce and Tableau Software are designed to support the use of analytics tools without advanced technical knowledge. Furthermore, SAS and IBM SPSS address the substantial challenges of managing data at the scale of the Internet (Baker & Siemens, 2013; Dean & Sanjay, 2008). Currently, the best platforms for harnessing the power of Big Data are open and flexible. They also blend the right technologies, tools and features to turn data compilation into data insight.

Big Data and analytics in higher education

Big Data is a knowledge system that is already changing the objects of knowledge and social theory in many fields while also having the potential to transform management decision-making theory (Boyd & Crawford, 2012). Big Data incorporates the emergent research field of learning analytics (Long & Siemen, 2011), which is already a growing area in education. However, research in learning analytics has largely been limited to examining indicators of individual student and class performance. Big Data brings new opportunities and challenges for institutions of higher education. Long and Siemen (2011) indicated that Big Data presents the most dramatic framework in efficiently utilising the vast array of data and ultimately shaping the future of higher education. The application of Big Data in higher education was also echoed by Wagner and Ice (2012), who noted that technological developments have certainly served as catalysts for the move towards the growth of analytics in higher education.

In the context of higher education, Big Data connotes the interpretation of a wide range of administrative and operational data gathered processes aimed at assessing institutional performance and progress in order to predict future performance and identify potential issues related to academic programming, research, teaching and learning (Hrabowski, Suess & Fritz, 2011a, 2011b; Picciano, 2012). Others indicated that to meet the demands of improved productivity, higher education has to bring the tool of analytics into the system. As an emerging field within education, a number of scholars have contended that Big Data framework is well positioned to address some of the key challenges currently facing higher education (see, eg, Siemens, 2011).

At this early stage much of the work on analytics within higher education is coming from interdisciplinary research, spanning the fields of Educational Technology, Statistics, Mathematics, Computer Science and Information Science. A core element of the current work on analytics in education is centred on data mining.

Big Data in higher education also covers database systems that store large quantities of longitudinal data on students' right down to very specific transactions and activities on learning and teaching. When students interact with learning technologies, they leave behind data trails that can reveal their sentiments, social connections, intentions and goals. Researchers can use such data to examine patterns of student performance over time—from one semester to another or from 1 year to another.

On a higher level, it could be argued that the added value of Big Data is the ability to identify useful data and turn it into usable information by identifying patterns and deviations from patterns. Long and Siemen (2011) indicated that Big Data is now well positioned to start addressing some of the key challenges currently facing higher education. An OECD (2013) report suggested that it may be the foundation on which higher education can reinvent both its business model and bring together the evidence to help make decisions about educational outcomes.

From an organisational learning perspective, it is well understood that institutional effectiveness and adaptation to change relies on the analysis of appropriate data (Rowley, 1998) and that today's technologies enable institutions to gain insights from data with previously unachievable levels of sophistication, speed and accuracy (Jacqueline, 2012). As technologies continue to penetrate all facets of higher education, valuable information is being generated by students, computer applications and systems (Hrabowski & Suess, 2010).

Furthermore, Big Data Analytics could be applied to examine student entry on a course assessment, discussion board entries, blog entries or wiki activity, which could generate thousands of transactions per student per course. These data would be collected in real or near real time as it is transacted and then analysed to suggest courses of action. As Siemens (2011) indicated that "[learning] analytics are a foundational tool for informed change in education" and provide evidence on which to form understanding and make informed (rather than instinctive) decisions.

Big Data can also address the challenges associated with finding information at the right time when data are dispersed across several unlinked different data systems in institutions. By identifying ways of aggregating data across systems, Big Data can help improve decision-making capability.

Context and conceptual framework

Big Data can influence higher education practice, from enhancing students experience to improved academic programming, to more effective evidence-based decision making, and to strategic response to changing global trends. Big Data promises to turn complex, often unstructured data into actionable information. Hilbert (2013) pointed out that Big Data delivers a cost-effective prospect to improve decision making. In order to organise the available literature and develop a research design to help frame set of approaches for investigation, Daniel and Butson (2013) proposed a conceptual framework to describe Big Data in higher education along four components (see Figure 3), utilising the framework as a way to describe and link siloed data systems (see Figure 4).

The University of Otago Technology Enhanced Analytics Framework (Figure 3) describes a wide range of administrative and operational data gathering processes aimed at assessing institutional performance and progress in order to predict future performance and identifies potential issues related to academic programming, research, teaching and learning.

Institutional analytics

Institutional analytics refers to a variety of operational data that can be analysed to help with effective decisions about making improvements at the institutional level. Institutional analytics include assessment policy analytics, instructional analytics, and structural analytics. Institutional analytics make use of reports, data warehouses and data dashboards that provide an institution with the capability to make timely data-driven decisions across all departments and divisions.

Information technology (IT) analytics

IT analytics covers usage and performance data that helps with monitoring required for developing or deploying technology, developing data standards, tools, processes, organisational synergies and policies. IT analytics aim at integrating data from a variety of systems—student information, learning management and alumni systems, as well as systems managing learning experiences outside the classroom.

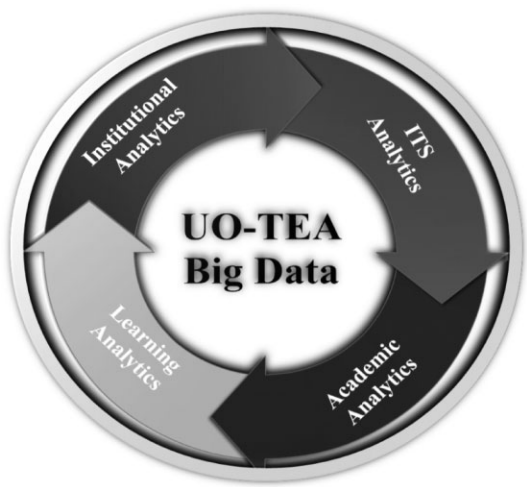

Figure 3: Conceptual framework

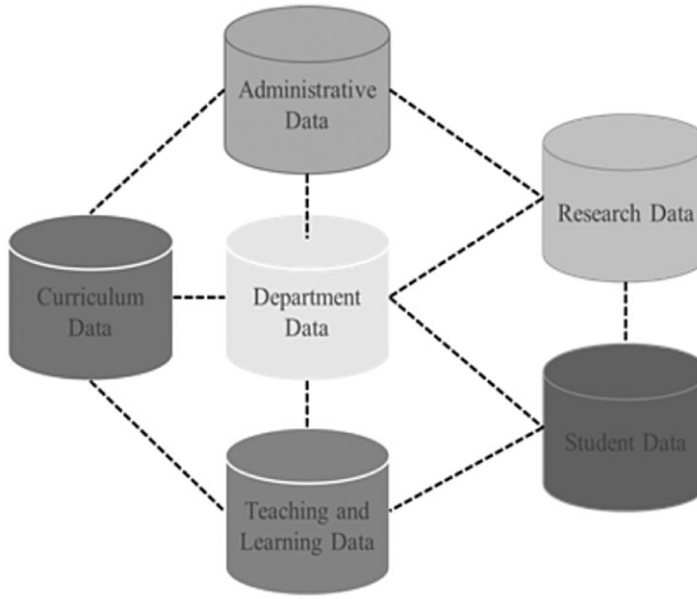

Figure 4: Ideal data framework

Results of IT analytics are used to develop rigorous data modelling and analysis to reveal the obstacles to student access and usability and to evaluate any attempts at intervention. Hrabowski and Suess (2010) reported that with analytics, IT systems can help by refining the associated business processes to collect critical data that might not have been collected institutionally and by showing how data in separate systems can become very useful when captured and correlated.

Academic/programme analytics

Academic analytics will be an essential component of the future. It encapsulates all the activities in higher education affecting administration, research, resource allocation and management (Tulasi, 2013). Academic analytics provides overall information about what is happening in a specific programme and how to address performance challenges. Academic analytics reflects the role of data analysis at an institutional level, whereas learning analytics centres on the learning process (which includes analysing the relationship between learner, content, institution, and educator) (Long & Siemen, 2011).

Academic analytics combines large data sets with statistical techniques and predictive modelling to improve decision making. Academic analytics provide data that administrators can use to support the strategic decision-making process as well as a method for benchmarking in comparison with other institutions.

The goal of an academic analytics programme is also to help those charged with strategic planning in a learning environment to measure, collect, interpret, report and share data in an effective manner so that operational activities related to academic programming and student strengths and weaknesses can be identified and appropriately rectified.

Learning analytics

As higher education institutions adopt blended approaches to teaching, learning is happening more and more within online environments and platforms. The educational data mining community and learning modelling communities have already explored ways to track student behaviours, recording variables such as number of clicks and time spent on a page, and increas-

ingly more nuanced information such as resilience and retention of concepts. Inclusion of behaviour-specific data adds to an ever-growing repository of student-related information. Learning analytics is an emergent research area that intends to access and understand these data and adds a new dimension to the learning process.

Learning analytics is concerned with the measurement, collection, and analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs (Long & Siemen, 2011). More broadly, learning analytics software and techniques are commonly used for improving processes and workflows, measuring academic and institutional data and generally improving organisational effectiveness (Dyckhoff, Zielke, Bültmann, Chatti & Schroeder, 2012; Jones, 2012).

Although such usage is often referred to as learning analytics, it is more associated with 'academic analytics' (Goldstein & Katz, 2005). Learning analytics is undertaken more at the teaching and learning level of an institution and is largely concerned with improving learner success (Jones, 2012). Associated with learning analytics is the notion of teacher analytics. Xu and Recker (2012) extended learning analytics to include teaching analytics. In their work, they analysed teachers' online behaviours in the context of utilising digital library and online resources. They utilised educational data mining techniques to identify different groups of instructional architectures to determine diverse online behaviours.

Opportunities

With large volumes of student information, including enrollment, academic and disciplinary records, institutions of higher education have the data sets needed to benefit from a targeted analytics. Big Data and analytics in higher education can be transformative, altering the existing processes of administration, teaching, learning, academic work (Baer & Campbell, 2011), contributing to policy and practice outcomes and helping address contemporary challenges facing higher education.

Big Data can provide institutions of higher education the predictive tools they need to improve learning outcomes for individual students as well ways ensuring academic programmes are of high-quality standards. By designing programmes that collect data at every step of the students learning processes, universities can address student needs with customised modules, assignments, feedback and learning trees in the curriculum that will promote better and richer learning.

One of the ways higher education can utilise Big Data tools is to analyse the performance and skill level of individual students and create a personalised learning experiences that meet their specific learning pathways. When used effectively, Big Data can help institutions enhance learning experience and improving student performance across the board, reduce dropout rates and increase graduation numbers (Figure 5).

The key contribution of Big Data will depend on the application of three data models (descriptive, relational and predictive) and the utility of each to guide better decision making (Figure 6).

Descriptive analytics

Descriptive analytics aims at describing and analysing historical data collected on students, teaching, research, policies and other administrative processes. The goal is to identify patterns from samples to report on current trends—such as student enrollment, graduation rates and progressions into higher degrees.

Descriptive analytics also provides institutions of higher education with an opportunity to analyse transactional and interactional data about teaching learning and research to identify discernible trends and patterns that are likely to trigger important dialogue on current and future issues. Specifically, with descriptive analytics, institutions can investigate data within learning management systems by looking into the frequency of logins, page views, course completion

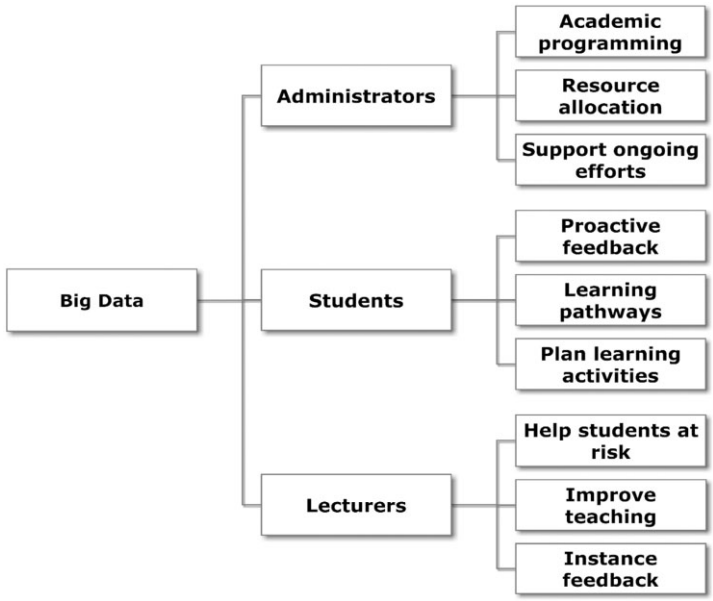

Figure 5: Key Big Data opportunities for three end-users in higher education

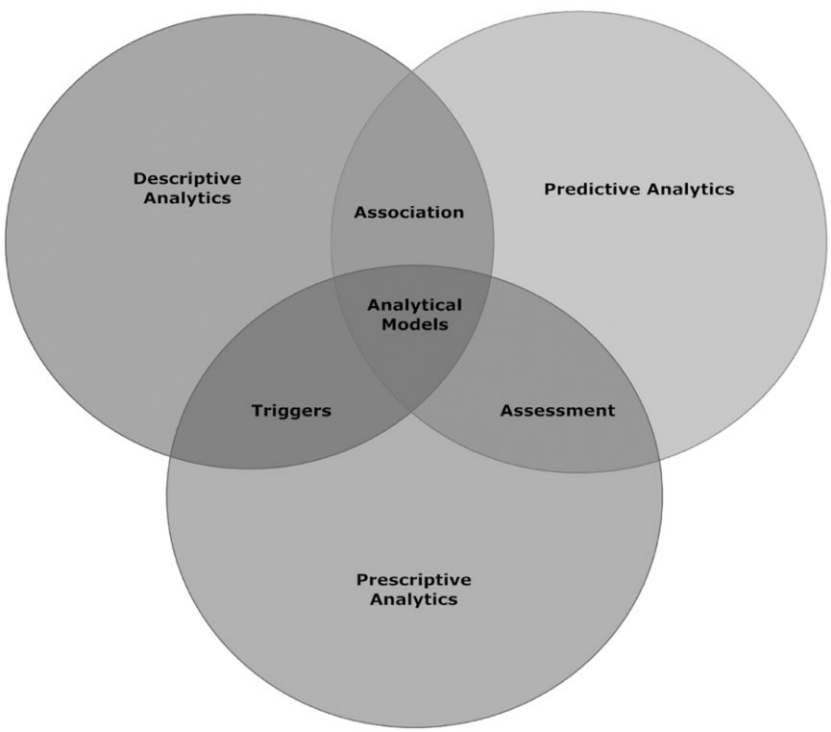

Figure 6: Three Big Data Analytical models in higher education

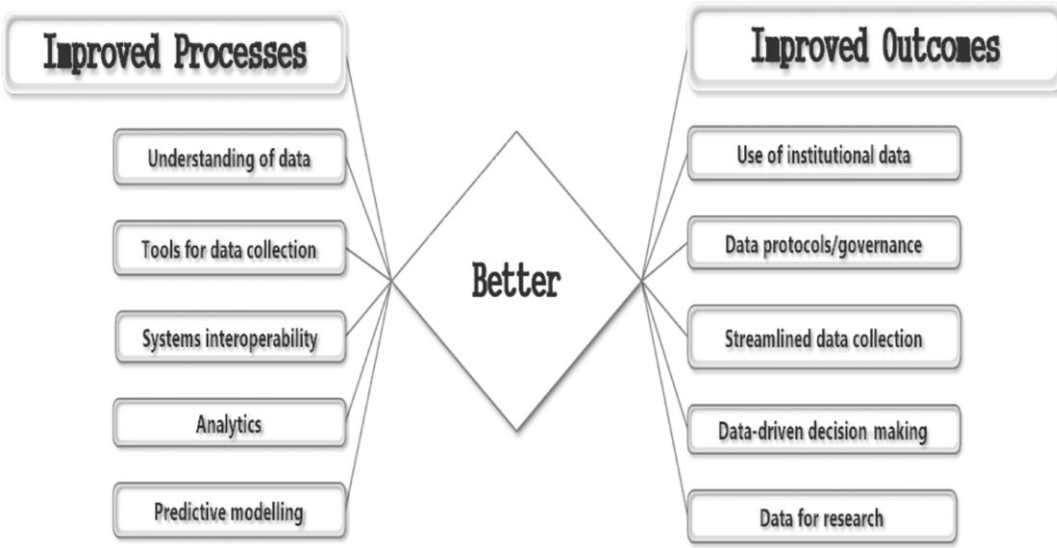

Figure 7: Big Data analytics outcomes

rates for a particular programme over time, students attributes of those who completed versus those struggling, which content is being visited many times, etc.

Predictive analytics

Predictive analytics can provide institutions with better decisions and actionable insights based on data. Predictive analytics aims at estimating likelihood of future events by looking into trends and identifying associations about related issues and identifying any risks or opportunities in the future. Predictive analytics could reveal hidden relationships in data that might not be apparent with descriptive models, such as demographics and completion rates.

It can also be used help to look at students who are exhibiting risk behaviours early in the semester that might result to dropping out or failing a course. It can help teachers look at predicted course completion rate for a particular and tools and content in the course are directly correlated to student success.

Prescriptive analytics

Prescriptive analytics helps institutions of higher education assess their current situation and make informed choices on alternative course of events based on valid and consistent predictions. It combines analytical outcomes from both descriptive and predictive models to look at assessing and determining new ways to operate to achieve desirable outcomes while balancing constraints. Basu (2013) indicated that prescriptive analytics enables decision makers to look into the future of their mission critical processes and see the opportunities (and issues) as well as presents the best course of action to take advantage of that foresight in a timely manner.

In summary, Big Data Analytics provides institutions of higher education to leverage existing data and collect missing data to help make better decisions, with various outcomes (see Figure 7).

In particular, Big Data helps institutions with the following performance and process outcomes:

Performance outcomes

- Better understanding of institutional data
- Better understanding of the requirements for effective data preparation for analytics

- A solid foundation for the utilisation of Big Data
- Improved standardised and streamlined data processes
- Consistent ways to effectively leverage data analytics for improved accuracy, deeper knowledge and real time decision making
- Better data-driven decision making and practice
- Foundation for hypothesis testing, web experimenting, scenario modelling, simulation, sensibility and data mining.

Process outcomes

- Better tools for collecting, processing, analysing and interpretation of data
- Better data system interoperability and system linking
- Enhanced data analytics and predictive modelling
- Better real-time rendering of analytics on students and instructors performances
- Reliable and comparable performance indicators and metrics within departments and divisions
- Better utilisation of historical institutional data to make informed decisions
- Better ability to develop and utilise "what if" scenarios for exploring data to predict possible outcomes.

Challenges of implementation

There are a number of anticipated challenges associated with the implementation of analytic techniques for Big Data in higher education. Some of these include challenges associated with getting users to accept Big Data as a conduit for adopting new processes and change management. There is also a tremendous cost associated with collecting, storing and developing algorithms to mine data, a process that tend to be time consuming and complex. Furthermore, most of institutional data systems are not interoperable, so aggregating administrative data, classroom and online data can pose additional challenges (Daniel & Butson, 2013).

Furthermore, data integration challenges are eminent, especially where data come in both structured and unstructured formats and needed to be integrated from disparate sources most of which are stored in systems managed by different departments. Additionally, data cleansing when performing integration of structured and unstructured data is likely to result to loss of data.

There are also challenges associated with quality of data collected and reported. Lack of standardised measures and indicators make inter(national) comparison difficult, as the quality of information generated from Big Data is totally dependent on the quality of data collected and the robustness of the measures or indicators used.

A recent U.S. Department of Education (2012) report suggested that the successful implementation of Big Data in higher institution would depend on collaborative initiatives between various departments in a given institution. For instance, the involvement of IT services departments in planning for data collection and use is deemed critical.

This is consistent with views that the value of Big Data will be based on the ability to co-create governing structures and delivery of more progressive and better policies and strategies currently used (Schleicher, 2013). Wagner and Ice (2012) also pointed out that by increasing collaborative ventures on Big Data initiatives help all groups take ownership of the challenge involving student performance and persistence. Dringus (2012) suggested that the practice of learning analytics should be transparent and flexible to make it accessible to educators (Dringus, 2012; Dyckhoff *et al*, 2012).

However, there is still a divide between those who know how to extract data and what data are available, and those who know what data are required and how it would best be used, all which make collaboration difficult. Furthermore, as Romero and Ventura (2010) note, analytics has

traditionally been difficult for non-specialists to generate (and generate in meaningful context), to visualise in compelling ways, or to understand, limiting their observability and decreasing their impact (Macfadyen & Dawson, 2012).

The importance of communicating these ideas is also acknowledged by Macfadyen and Dawson (2012), who pointed out that analytics have a negative or neutral impact on educational planning. They advocate delving into the socio-technical sphere to ensure analytics are presented to those involved in strategic positions in ways that have the power to motivate organisational adoption and cultural change.

Becker (2013) suggested three interactive components to be studied when collecting data for analytics: location, population and timing. Location is defined by where and how students are accessing the learning space, whereas population refers to the characteristics of the group of learners participating in the learning space. Timing can be defined by any unit, from second or minute to semester or year.

Big Data utilisation also raises issues around ethics of data collection in regard to quality of data, privacy, security and ownership. It also raises the question of an institutions responsibility for taking action based on the information available (Jones, 2012). Security and privacy issues pose additional challenge to implementation of Big Data in higher education.

Currently, techniques such as disaster recovery plans, strong password policy, firewalls, encryption and anti-virus software that reduce the risks of losing or manipulating Big Data are still being investigated. Furthermore, risks and security procedures for data protection and privacy are still lacking in many institutions of higher education. For instance, Slade and Prinsloo (2013) pointed out that while most higher education institutions seem to have policies to regulate and govern intellectual property, safeguard data privacy and regulate access to data, these policies might not be adequate to address contemporary challenges associated with Big Data in higher education.

Dringus (2012) for example argue that bringing transparency to learning analytics as a practice could be used to help deter any potentially wrongful use of data. As the amount of data available for use is ever-increasing, the benefits will come from good learning management, reliable data warehousing and management, flexible and transparent data mining and extraction, and accurate and responsible reporting.

Summary and future directions

Dramatic advances in data capture, processing power, data transmission and storage capabilities are enabling many organisations to integrate their various databases into data warehouses. In the age of abundant data, higher education institutions similar to business, government or healthcare institutions share some of the same reasons for adopting analytics, especially in the areas of financial efficiency; expanding local and global impact; addressing new funding models during a changing economic climate; and responding to the demands for greater accountability.

Within institutions of higher education, data are growing, but most of it is scattered across desktops, departments and come in various formats, making it difficult to retrieve or consolidate. To effectively utilise these data, the ability to analyse diverse information sets is needed, regardless of where they originate, and consolidating data stored in silos within institutions, managing and governing the data while securing sensitive information across databases, is a key requirement for implementation of Big Data in higher education (Daniel & Butson, 2013).

Consolidated data from various sources across an institution provide a better foundation for making better decisions related to key business and technical needs, reducing redundancies and waste of precious time retrieving data from multiple sources. In other words, data warehousing can be an effective approach to unlock the value of Big Data in higher education. Although combining data sets from across a variety of unconnected systems can be extremely difficult, it

offers better comprehensive insights that inevitably lead to improved capabilities of predictive modelling. Dringus (2012) suggested that one way of overcoming these problems is to increase institutional transparency by clearly demonstrating the changes that analytics can help to achieve.

Analytics also has the potential to help learners and instructors recognise danger signs before threats to learning success materialise (Wagner & Ice, 2012); however, wide institutional acceptance of analytics requires a clear institutional strategy and the usability of analytics software packages (Ali, Adasi, Gasevic, Jovanovic & Hatala, 2013).

Furthermore, data warehousing implements standardised data formats. Each department will produce results that are standardised with all the other departments, resulting in more accurate data representation. Lastly, a data warehouse can store large amounts of historical data that can be readily for experimentation and to analyse different time periods and trends in order to make future predictions.

Big Data can help provide insights to support student's learning needs. For instance, learning analytics as a fundamental component of Big Data in higher education provide researchers with opportunities to carry out real-time analysis of learning activities. By performing retrospective analysis of student data, predictive models can be created to examine students at risk and provide appropriate intervention, hence enabling instructors to adapt their teaching or initiate tutoring, tailored assignments and continuous assessment.

Despite the substantial uncertainties, the continuing growth of learning analytics means that we need to consider not only the vast opportunities offered for better and more effective decision making in higher education (Oblinger, 2012) but also explore the ethical challenges in institutionalising learning analytics as a means to drive and shape student support (Slade & Prinsloo, 2013).

Work is underway exploring data management and governance structures associated with Big Data in higher education. This research is also looking developing conceptual and theoretical underpinnings of Big Data and analytics in higher education, as well as developing key performance indicators, metrics and methods for capturing, processing and visualising data. In addition, a set of diagnostic tools and an integrated technology-enhanced data analytic framework and ultimately a data warehouse for Big Data Analytics are being pursued.

Future work will involve identifying and establishing policies that specify who is accountable for various portions or aspects of institutional data and information, including its accuracy, accessibility, consistency, completeness and maintenance.

This research will also be looking at defining processes concerning how data and information are stored, archived, backed up and protected, as well as developing standards and procedures that define how the data and information are used by authorised personnel and implement a set of audit and control procedures to ensure ongoing compliance with governmental regulations and industrial standards.

References

Agrawal, R., Imielinski, T. & Swami, A. (1993). Mining Associations Between Sets of Items in Massive Databases. Proceedings of the ACM-SIGMOD 1993 Int'l Conference on Management of Data, Washington D.C., May 1993.

Ali, L., Adasi, M., Gasevic, D., Jovanovic, J. & Hatala, M. (2013). Factors influencing beliefs for adoption of a learning analytics tool: an empirical study. *Computers & Education*, *62*, 130–148.

Baepler, P. & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, *4*, 2, 1–9. Retrieved September 17, 2014, from http://digitalcommons.georgiasouthern.edu/ij-sotl/vol4/iss2/17

Baer, L. & Campbell, J. (2011). Game changers. *EDUCAUSE*. Retrieved March 24, 2014, from http://net.educause.edu/ir/library/pdf/pub72034.pdf

Baker, R. S. J. D. & Siemens, G. (2013). *Educational data mining and learning analytics*, Cambridge: Cambridge University.

Baker, R. S. J. D. (2013). *Learning, schooling, and data analytics. Handbook on innovations in learning for states, districts, and schools* (pp. 179–190). Philadelphia, PA: Center on Innovations in Learning.

Baker, R. S. J. D. & Inventado, P. S. (in press). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds), *Learning analytics: from research to practice*. Berlin, Germany: Springer.

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

Basu, A. (2013). Five pillars of prescriptive analytics success. Analytics, pp. 8–12, March/April Issue [ANALYTICS-MAGAZINE.ORG].

Becker, B. (2013). Learning analytics: insights into the natural learning behavior of our students. *Behavioral & Social Sciences Librarian*, *32*, 1, 63–67. doi: 10.1080/01639269.2013.751804.

Borgman, C., Abelson, H., Dirks, L., Johnson, R., Koedinger, K., Linn, M. *et al* (2008). *Fostering learning in the networked world: The cyberlearning opportunity and challenge, a 21st century agenda for the national science foundation. Arlington, VA: National Science Foundation*. Retrieved April 18, 2014, from http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf

Boyd, D. & Crawford, K. (2012). Critical questions for Big Data. *Communication & Society*, *15*, 5, 662–679. doi: 10.1080/1369118X.2012.678878.

Choudhury, S., Hobbs, B., Lorie, M. & Flores, N. (2002). A framework for evaluating digital library services. *D-Lib Magazine*, 8, 7/8. Retrieved October 30, 2014, from http://www.dlib.org/dlib/july02/choudhury/07choudhury.html.

Clarke, J., Nelson, K. & Stoodley, I. (2013). The place of higher education institutions in assessing student engagement, success and retention: A maturity model to guide practice. In S. Frielick, N. Buissink-Smith, P. Wyse, J. Billot, J. Hallas & E. Whitehead (Eds), *Proeedings of Research and Development in Higher Education Conference: The Place of Learning and Teaching* Vol. 36 (pp. 91–101). Auckland, New Zealand, 1–4 July 2013.

Daniel, B. K. & Butson, R. (2013). Technology enhanced analytics (TEA) in higher education, Proceedings of the International Conference on Educational Technologies, 29 Novemebr–1 December, 2013, Kuala Lumpur, Malaysia, pp. 89–96.

Dean, J. & Sanjay, G. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*, 1, 107–113.

Douglas, L. (2001). 3D data management: controlling data volume, velocity and variety. *Gartner Report*. Retrieved December 30, 2013, from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Dringus, L. P. (2012). Learning analytics considered harmful. *Journal of Asynchronous Learning Networks*, *16*, 3, 87–100.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A. & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology & Society*, *15*, 3, 58–76.

Economist Report (2008). *The future of higher education: how technology will shape learning. A report from the Economist Intelligence Unit*. Retrieved March 24, 2014, from http://www.nmc.org/pdf/Future-of-Higher-Ed-(NMC).pdf

Goldstein, P. J. & Katz, R. N. (2005). *Academic Analytics: The Uses of Management Information and Technology in Higher Education*, ECAR Research Study Volume 8. Retrieved March 24, 2014, from http://www.educause.edu/ers0508

Hazelkorn, E. (2007). The impact of league tables and ranking systems on higher education decision making. *Higher education management and policy*, *19*, 2, 1–24.

Hilbert, M. (2013). *Big Data for Development: From Information- to Knowledge Societies (January 15, 2013)*. Retrieved October 30, 2014, from http://ssrn.com/abstract=2205145 or http://dx.doi.org/10.2139/ssrn.2205145

Hrabowski, F. A. III & Suess, J. (2010). Reclaiming the lead: higher education's future and implications for technology. *EDUCAUSE Review*, *45*, 6 (November/December 2010). Retrieved October 30, 2014, from http://www.educause.edu/library/ERM1068

Hrabowski, F. A. III, Suess, J. & Fritz, J. (2011a). *Analytics in institutional transformation*. EDUCAUSEREVIEW, pp. 15–28. Retrieved March 24, 2014, from https://net.educause.edu/ir/library/pdf/ERM1150.pdf

Hrabowski, F. A. III, Suess, J. & Fritz, J. (2011b). Assessment and analytics in institutional transformation. Assessment and analytics in institutional transformation (EDUCAUSE review). *EDUCAUSE Review*, *46*, 5 (September/October 2011). Retrieved August 8, 2013, from http://www.educause.edu/ero/article/assessment-and-analytics-institutional-transformation

Jacqueline, B. (2012). *Analytics in Higher Education: Benefits, Barriers, Programs and Recommendations*. Retrieved March 24, October 30, 2014, from http://net.educause.edu/ir/library/pdf/ERS1207/ers1207.pdf

Jones, S. (2012). Technology review: the possibilities of learning analytics to improve learner-centered decision-making. *Community College Enterprise*, *18*, 1, 89–92.

Kline, P. (1993). An Easy Guide to Factor Analysis.

Leydesdorff, L. & Etzkowitz, H. (2001). The transformation of university–industry–government relations. *Electronic Journal of Sociology*, 5, 4, 1–17.

Long, P. & Siemen, G. (2011). Penetrating the fog: analytics in learning and education. *EDUCAUSE Review*, *46*, 5, 30–40.

Luan, J. (2002). Data mining and its applications in higher education. In A. Serban & J. Luan (Eds), *Knowledge management: building a competitive advantage in higher education* (pp. 17–36). PA: Josey-Bass.

Macfadyen, L. P. & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, *15*, 3, 149–163.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. *et al* (2011). *Big data: the next frontier for innovation, competition, and productivity*. Retrieved October 7, October 30, 2014, from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

Mayer, M. (2009). *Innovation at Google: The physics of data [PARC forum]*. Retrieved 11 August, 2009, from http://www.slideshare.net/PARCInc/innovation-at-google-the-physics-of-data

Minaei-Bidgoli, B., Kashy, D., Kortmeyer, G. & Punch, W. (2003). Predicting student performance: an application of data mining methods with an educational web-based system. In: Frontiers in education, 2003. FIE 2003 33rd Annual, pp T2A 13–18.

Mok, K. H. (2005). Fostering entrepreneurship: changing role of government and higher education governance in Hong Kong. *Research Policy*, *34*, 4, 537–554.

Oblinger, D. G. (2012). Let's talk analytics. *EDUCAUSE Review*, *July/August*, 10–13.

OECD (2013). *OECD Report: The State of Higher Education 2013*. Retrieved March 24, 2014, from http://www.oecd.org/edu/imhe/thestateofhighereducation2013.htm

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. R. & 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.

Rowley, J. (1998). Creating a learning organisation in higher education. *Industrial and Commercial Training*, *30*, 1, 16–19.

Schleicher, A. (2013). *Big Data and PISA*. Retrieved August 4, 2013, from http://oecdeducationtoday.blogspot.co.nz/2013/07/big-data-and-pisa.html?m=1

Siemens, G. (2011). *How data and analytics can improve education, July 2011*. Retrieved August 8, October 30, 2014, from http://radar.oreilly.com/2011/07/education-data-analytics-learning.html

Slade, S. & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, *57*, 10, 1509–1528.

Thornton, G. (2013). *The state of higher education in 2013. Pressures, changes and new priorities*. Retrieved March 24, 2014, from http://www.grantthornton.com/staticfiles/GTCom/Not-for-profit%20organizations/The%20state%20of%20higher%20education%20in%202013.pdf

Tulasi, B. (2013). Significance of Big Data and analytics in higher education. *International Journal of Computer Applications*, *68*, 14, 23–25.

U.S. Department of Education (2012). Office of Educational Technology, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Washington, D.C.

Wagner, E. & Ice, P. (2012). Data changes everything: delivering on the promise of learning analytics in higher education. *EDUCAUSE Review*, *July/August*, 33–42.

Xu, B. & Recker, M. (2012). Teaching analytics: a clustering and triangulation study of digital library user data. *Educational Technology & Society*, *15*, 3, 103–115.

Yang, L. (2013). Big Data Analytics: What Is the Big Deal? Retrieved December 30, 2013, from http://knowledge.ckgsb.edu.cn/2013/12/30/technology/big-data-analyticswhats-big-deal/