Learning Analytics Ethical Issues and Dilemmas

Article *in* American Behavioral Scientist · October 2013

DOI: 10.1177/0002764213479366

2 authors:

Sharon Slade
The Open University (UK)

Paul Prinsloo
University of South Africa

American Behavioral Scientist

http://abs.sagepub.com/

Learning Analytics: Ethical Issues and Dilemmas

Sharon Slade and Paul Prinsloo

American Behavioral Scientist 2013 57: 1510 originally published online 4 March 2013
DOI: 10.1177/0002764213479366

The online version of this article can be found at:
http://abs.sagepub.com/content/57/10/1510

Published by:
SAGE
http://www.sagepublications.com

Article

American Behavioral Scientist
57(10) 1510–1529

DOI: 10.1177/0002764213479366
abs.sagepub.com

Learning Analytics: Ethical Issues and Dilemmas

Sharon Slade[1] and Paul Prinsloo[2]

Abstract

The field of learning analytics has the potential to enable higher education institutions to increase their understanding of their students' learning needs and to use that understanding to positively influence student learning and progression. Analysis of data relating to students and their engagement with their learning is the foundation of this process. There is an inherent assumption linked to learning analytics that knowledge of a learner's behavior is advantageous for the individual, instructor, and educational provider. It seems intuitively obvious that a greater understanding of a student cohort and the learning designs and interventions they best respond to would benefit students and, in turn, the institution's retention and success rate. Yet collection of data and their use face a number of ethical challenges, including location and interpretation of data; informed consent, privacy, and deidentification of data; and classification and management of data. Approaches taken to understand the opportunities and ethical challenges of learning analytics necessarily depend on many ideological assumptions and epistemologies. This article proposes a *sociocritical* perspective on the use of learning analytics. Such an approach highlights the role of power, the impact of surveillance, the need for transparency, and an acknowledgment that student identity is a transient, temporal, and context-bound construct. Each of these affects the scope and definition of learning analytics' ethical use. We propose six principles as a framework for considerations to guide higher education institutions to address ethical issues in learning analytics and challenges in context-dependent and appropriate ways.

Keywords

ethics, framework, higher education, learning analytics

[1]The Open University, Oxford, UK
[2]University of South Africa, Unisa, South Africa

Corresponding author:

Sharon Slade, Faculty of Business and Law, The Open University, Foxcombe Hall, Boars Hill, Oxford, UK.
Email: sharon.slade@open.ac.uk

Learning analytics is cited as one of the key emerging trends in higher education (Booth, 2012; Johnson, Adams, & Cummins, 2012). There remain, however, a number of issues in need of further research and reflection (Siemens, 2012), not least of which is consideration of its ethical use and practices. In the field of learning analytics, discussions around the ethical implications of increasing an institution's scrutiny of student data typically relate to ownership of that data and to student privacy issues. Although many authors in the field refer to ethical issues, there are few integrated and coherent attempts to map ethical concerns and challenges pertaining to the use of learning analytics in higher education.

Approaches taken to understand the opportunities and ethical challenges of learning analytics necessarily depend on a range of ideological assumptions and epistemologies. For example, if we approach learning analytics from the perspective of resource optimization in the context of the commoditization of higher education (Hall & Stahl, 2012), the ethical issues appear different from those resulting from a sociocritical perspective. In addition, learning analytics has evolved from a range of research areas such as social network analysis, latent semantic analysis, and dispositions analysis (Ferguson, 2012). Each of these domains has its own, often overlapping, ethical guidelines and codes of conduct that address similar concerns such as the ownership of data, privacy, consumer or patient consent, and so on.

For this article, we situate learning analytics within an understanding of power relations among learners, higher education institutions, and other stakeholders (e.g., regulatory and funding frameworks). Such power relations can be considered from the perspective of Foucault's Panopticon, where structural design allows a central authority to oversee all activity. In the case of learning analytics, such oversight or surveillance is often available only to the institution, course designers, and faculty, and not to the student (Land & Bayne, 2005).

This article references existing research on learning analytics, adding an integrated overview of different ethical issues from a sociocritical perspective. A sociocritical perspective entails being critically aware of the way our cultural, political, social, physical, and economic contexts and power relationships shape our responses to the ethical dilemmas and issues in learning analytics (see, e.g., Apple, 2004). Ethical issues for learning analytics fall into the following broad, often overlapping categories:

1. The location and interpretation of data
2. Informed consent, privacy, and the deidentification of data
3. The management, classification, and storage of data

Such ethical issues are not unique to education, and similar debates relating to the use of data to inform decisions and interventions are also found in the health sector (Cooper, 2009; Dolan, 2008; Snaedal, 2002), human resource management (Cokins, 2009), talent management (Davenport, Harris, & Shapiro, 2010), homeland security (Seifert, 2004), and biopolitics (Dillon & Loboguerrero, 2008). Of specific concern here are the implications of viewing learning analytics as moral practice, recognizing students as participatory agents with developmental and temporal identities and learning trajectories and the need for reciprocal transparency. Learning analytics as moral practice functions as a counternarrative to using student data in service of neoliberal consumer-driven market ideologies (see, e.g., Giroux, 2003).

We conclude this article by proposing a number of grounding principles and considerations from which context-specific and context-appropriate guidelines can be developed.

Perspectives on Ethical Issues in Learning Analytics

Published literature on the ethical considerations in learning analytics tends to focus on issues such as the historical development of research ethics in Internet research, the benefits of learning analytics for a range of stakeholders, and issues of privacy, informed consent, and access to data sets. Given that many of these overlap, the following review of literature is structured to highlight systematically a range of different ethical issues for learning analytics in higher education. We aim also to highlight a number of issues that either have not yet been considered within the context of learning analytics or have not been considered fully.

A Working Definition of Learning Analytics

Oblinger (2012) differentiates between learning analytics and approaches including business intelligence and academic analytics, defining learning analytics as focusing on "students and their learning behaviors, gathering data from course management and student information systems in order to improve student success" (p. 11). For this article, we define learning analytics as the collection, analysis, use, and appropriate dissemination of student-generated, actionable data with the purpose of creating appropriate cognitive, administrative, and effective support for learners.

An Overview of the Purposes and Collection of Educational Data

Ethical challenges and issues in the use of educational data can be usefully viewed in the context of the history of Internet research ethics and against the backdrop of the development of research ethics after cases such as the Tuskegee experiment (Lombardo & Dorr, 2006), the release of the Nuremberg Code in 1947, and the World Medical Association's Declaration of Helsinki adopted in 1964. Throughout there has emerged an attempt to find a balance between "individual harms and greater scientific knowledge" (Buchanan, 2011, p. 84). The advent of the Internet exposed new areas for concern, often lying outside traditional boundaries and guidelines for ethical research. As a way to provide guidance for the complexities of conducting research on Internet populations and data, Internet research ethics emerged in the early 1990s. This was followed in 2000 by the formation of the Ethics Working Group by the Association of Internet Researchers.

Higher education institutions have always analyzed data to some extent, but the sheer volume of data continues to rise along with institutions' computational capacity, the prevalence of visualization tools, and the increasing demand for the exploitation of data. As a result, there are a growing number of ethical issues regarding the collection and analyses of educational data, issues that include greater understanding and transparency regarding the "purposes for which data is being collected and how sensitive data will be handled" (Oblinger, 2012, p. 12). Legal frameworks such as the U.S. Family Educational Rights and Privacy Act focus largely on how information is used outside an institution rather than on its use within the institution, and Oblinger (2012) argues that it is crucial that institutions inform students "what information about them will be used for what purposes, by whom, and for what benefit" (p. 12).

Subotzky and Prinsloo (2011) suggest that there is a need for a reciprocal sharing of appropriate and actionable knowledge between students and the delivering institution. Such knowledge of students may facilitate the offering of just-in-time and customized support, allowing students to make more informed choices and act accordingly.

The Educational Purpose of Learning Analytics

In stark contrast to the advances of data use in other fields (e.g., patient information in health care), higher education "has traditionally been inefficient in its data use, often operating with substantial delays in analyzing readily evident data and feedback" (Long & Siemens, 2011, p. 32). Despite the probable advantages of using learning analytics to measure, collect, analyze, and report "data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long & Siemens, 2011, p. 34), there remain a number of ethical challenges and issues affecting its optimization in higher education.

Although educational data serve a number of purposes (e.g., for reporting on student success and study subsidies), Booth (2012) emphasizes that learning analytics also has the potential to serve learning. A learning analytics approach may make education both personal and relevant and allow students to retain their own identities within the bigger system. Optimal use of student-generated data may result in institutions having an improved comprehension of the lifeworlds and choices of students, allowing both institution and students to make better and informed choices and respond faster to actionable and identified needs (Oblinger, 2012).

Several authors (Bienkowski, Feng, & Means, 2012; Campbell, DeBlois, & Oblinger, 2007) refer to the obligation that institutions have to act on knowledge gained through analytics. It is fair to say that there may also be instances where institutions decide not to act on data. It might be argued that the gains offered by responding to student cohorts with a certain set of shared characteristics or behaviors are of minor benefit or are less beneficial than making equivalent or lower investments of resources elsewhere. The recent EDUCAUSE Center for Applied Research survey (Bichsel, 2012, p. 13) confirms that the greatest concern relating to the growing use of learning analytics expressed by a range of members is the financial cost of implementation, rather than issues of privacy or misuse of data.

The decision to act or not to act and the costs and ethical considerations of either have to be considered in the specific context of application. Not all data harvested will necessarily involve or trigger a response on the part of the institution. This raises the point that although not all data will be actionable, they may increase institutions' understanding of student success and retention.

At some point, all institutions supporting student learning must decide what their main purpose really is: to maximize the number of students reaching graduation, to improve the completion rates of students who may be regarded as disadvantaged in some way, or perhaps to simply maximize profits. The ways in which learning analytics is applied by an institution will vary in accordance with which of these is deemed to be its primary concern. Furthermore, and perhaps more important for the institution's ability to maintain positive relations with its students, the ways in which students perceive the use of such surveillance will also vary in accordance with their own understanding of the institution's purpose and motivation. The management of students' understanding and perceptions is therefore a major priority for any institution that seeks to embed learning analytics into its standard operations.

Amid the emphasis on the role of data and analyses for reporting on student success, retention, and throughput, it is crucial to remember that learning analytics has huge potential to primarily serve *learning* (Kruse & Pongsajapan, 2012). When institutions emphasize the analysis and use of data primarily for reporting purposes, there is a danger of seeing students as passive producers of data, resulting in learning analytics used as "intrusive advising" (Kruse & Pongsajapan, 2012, p. 2). If the intention is to use data collected from students for other purposes, there is a responsibility on the part of the institution to make that known.

Power and Surveillance in Learning Analytics

Owing to the inherently unequal power relations in the use of data generated by students in their learning journey, we propose a sociocritical framework that allows us to address a range of ethical questions such as levels of visibility, aggregation, and surveillance. Although online surveillance and the commercialization possibilities of online data are themselves under increasing scrutiny (e.g., Andrejevic, 2011; Livingstone, 2005), the ease with which we now share data on social networking sites may suggest an increasing insouciance or a less guarded approach to privacy (e.g., Adams & Van Manen, 2006). Dawson (2006), for example, found that students altered their online behaviors (e.g., range of topics discussed and writing style) when aware of institutional surveillance. Albrechtslund (2008, para. 46) explores the notion of participatory surveillance, focusing on surveillance as "mutual, horizontal practice" as well as the social and "playful aspects" of surveillance (also see Knox, 2010b; Lyon, 2007; Varvel, Montague, & Estabrook, 2007). Knox (2010b) provides a very useful typology of surveillance, highlighting, inter alia, the difference between surveillance and monitoring, automation and visibility, and various aspects of rhizomatic and predictive surveillance. Rhizomatic surveillance highlights the dynamic, multidirectional flow of the act of surveillance in a synopticon, where the many can watch the few. The synopticon and Panopticon function concurrently and interact (Knox, 2010b).

The increasing surveillance in teaching and learning environments also affects the work and identities of tutors, faculty, and administrators, disrupting existing power relations and instituting new roles and responsibilities (e.g., Knox, 2010a).

The following sections discuss a range of ethical issues grouped within three broad, overlapping categories:

1. The location and interpretation of data
2. Informed consent, privacy, and the deidentification of data
3. The management, classification, and storage of data

1. The Location and Interpretation of Data

It is now the case that "significant amounts of learner activity take place externally [to the institution] . . . records are distributed across a variety of different sites with different standards, owners and levels of access" (Ferguson, 2012, para. 6). This flags the difficulties associated with attempting to enforce a single set of guidelines relating to ethical use across such a range of sites, each with its own data protection standards, for instance.

In addition, there are questions around the nature and interpretation of digital data as fully representative of a particular student (cohort). Correlations between different variables may be assumed when dealing with missing and incomplete data around usage of the institution's learning management system (LMS; Whitmer, Fernandes, & Allen, 2012). Such assumptions may be influenced by the analyst's own perspectives and result in subconsciously biased interpretations. The distributive nature of networks and the inability to track activity outside of an institution's internal systems also affect the ability to get a holistic picture of students' lifeworlds. Not only do we not have all the data, a lot of the data that we do have require "extensive filtering to transform the 'data exhaust' in the LMS log file into educationally relevant information" (Whitmer et al., 2012, para. 22).

There are implications, too, of ineffective and misdirected interventions resulting from faulty learning diagnoses that might result in "inefficiency, resentment, and demotivation" (Kruse & Pongsajapan, 2012, p. 3). In addition, systematized modeling of behaviors, which necessarily involves making assumptions (e.g., regarding the permanency of students' learning contexts), can determine *and limit* how institutions behave toward and react to their students, both as individuals and as members of a number of different cohorts. Ess, Buchanan, and Markham (2012) concur that there is a need to consider the individual within what may be a very large and depersonalized data set, even if that individual is not recognizable. Actions influenced by a cohort of which a student is a single part may still adversely affect that student's options. In his recent article, Harvey (2012, para. 9) warns that "this process portends a reification of identities, with support allocated by association rather than individual need."

Given the wide range of information that may be included in such models, there is a recognized danger of potential bias and oversimplification (Bienkowski et al., 2012; Campbell et al., 2007; May, 2011). In accepting the inevitability of this, should we also question the rights of the student to remain an individual and whether it is appropriate for students to have an awareness of the labels attached to them? Are there some labels that should be prohibited? As students become more aware of the implications of such labeling, the opportunity to opt out or to actively misrepresent certain characteristics to avoid labeling can diminish the validity of the remaining data set. Many institutions are employing learning analytics to nudge students toward study choices or to adopt support strategies that are assumed to offer greater potential for success (Parry, 2012), but what is the obligation for the *student* to either accept explicit guidance or seek support that may be in conflict with his or her own preferences or study goals (Ferguson, 2012)? There is a risk of a "return to behaviorism as a learning theory if we confine analytics to behavioral data" (Long & Siemens, 2011, pp. 36-38).

2. Informed Consent, Privacy, and the Deidentification of Data

Although students are increasingly aware of the growing prevalence of data mining to monitor and influence buying behavior, it is not clear that they are equally aware of the extent to which this occurs within an educational setting. Epling, Timmons, and Wharrad (2003) discuss issues around the acceptability of student surveillance and debate who the real beneficiaries are. Use of data for noneducational purposes is flagged explicitly by Campbell et al. (2007), referring to the use of student-related data for fund-raising, for example. Wel and Royakkers (2004) discuss the ethics of tracking and analyzing student data without their explicit knowledge. Of interest, Land and Bayne (2005) discuss the broad acceptance of student surveillance and cite studies in which they record that the concept of logging educational activities seems to be quite acceptable to students. The notion of online privacy as a social norm is increasingly questioned (Arrington, 2010; Coll, Glassey, & Balleys, 2011).

Considering the general concern regarding surveillance and its impact on student and faculty privacy, Petersen (2012) points to the importance of the deidentification of data *before* the data are made available for institutional use, including the option to "retain unique identifiers for individuals in the data set, without identifying the actual identity of the individuals" (p. 48). This latter point addresses the need to provide interventions for groups of students based on their characteristics or behaviors while ensuring their anonymity within the larger data set.

3. The Classification and Management of Data

Petersen (2012) proposes a holistic approach to transparent data management, including a need for a "comprehensive data-governance structure to address all the types of data used in various situations," addressing the need to create "a data-classification system that sorts data into categories that identify the necessary level of protection" (pp. 46-47). He suggests a need to appoint data stewards to oversee standards and controls and set policies relating to, for example, data access, and data custodians to ensure adherence to policy and procedures without the power to determine who can and who cannot access data sets. Although Petersen's comments deal specifically with general data management, learning analytics might also benefit from such an approach.

With regard to the importance of trust in monitoring and surveillance (e.g., Adams & Van Manen, 2006; Knox, 2010a; Livingstone, 2005), we agree that the classification of data, as proposed by Petersen (2012), is an essential element in ensuring that appropriate access to different types of data will be regulated.

Integral to contemplating the ethical challenges in learning analytics is a consideration of the impact of the tools used. Wagner and Ice (2012) explore the relevance of pattern recognition and business intelligence techniques in the evolving learning analytics landscape that provide scope for increased success by guiding stakeholders to "recognize the proverbial right place and right time" (p. 34). Although pattern recognition has huge potential for delivering custom-made and just-in-time support to students, there is a danger, as highlighted by Pariser (2011), that pattern recognition can result in keeping individuals prisoner to past choices. Pariser suggests that the use of personalized filters hints of "autopropaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar," and that "knowing what kinds of appeals specific people respond to gives you the power to manipulate them on an individual basis" (p. 121). As we propose in the later section on ethical considerations, the algorithms used by institutions invariably reflect and perpetuate current biases and prejudices. The dynamic nature of student identity necessitates that we take reasonable care to allow students to act outside of imposed algorithms and models.

Student Identity as Transient Construct

Although the classification of data is the basis for determining access to different categories of data as well as determining appropriate institutional responses to different types of categories by a range of institutional stakeholders (including students themselves), it is crucial that the analysis of data in learning analytics keeps in mind the temporality of harvested data and the fact that harvested data allow us only a view of a person at a particular time and place. Although categorizing data is necessary, categorizing students and faculty based on historical data is, at least currently, error prone and incomplete. Institutions should also recognize the plurality of student identity. Sen (2006) suggests that we should recognize identities as "robustly plural, and that the importance of one identity need not obliterate the importance of others" (p. 19). Students, as agents, make choices—"explicitly or by implication—about what relative importance to attach, in a particular context, to the divergent loyalties and priorities that may compete for precedence" (p. 19; also see Brah, 1996).

Toward an Ethical Framework

Although the above literature review provided an overview on a range of ethical issues in learning analytics, we now turn to providing an integrated, sociocritical ethical framework and principles for learning analytics and discussion of a number of considerations that follow these principles.

The review left us with this question: How do we address both the potential of learning analytics to serve learning and the associated ethical challenges? One approach might be the formulation of institutional codes of conduct. Bienkowski et al. (2012), for example, cite work done in preparation of the U.S. Family Educational Rights and Privacy Act, which clarifies access to data sets (e.g., access primarily for research, accountability, or institutional improvement) set against the need to maintain student privacy. In Australia, Nelson and Creagh (2012) have begun work on a Good Practice Guide for using learning analytics, which details the needs of key stakeholders and the implications of a set of proposed rights for each. The viability of implementing such codes of conduct on a large scale and how these might address the use of all data in an online environment in which the data and their applications have the potential to increase and evolve should therefore be a priority in the debate of learning analytics and its ethical implications. Land and Bayne (2005) propose that institutional codes of conduct should cover informed consent, the purpose and extent of data tracking, the transparency of data held, ownership, and the boundaries of data usage.

Petersen (2012, p. 48) proposes adherence to the principles found in the U.S. Federal Trade Commission's Fair Information Practice Principles, which cover the elements of informed consent, allowing different options regarding the use of data, individuals' right to check the accuracy and completeness of information, preventing unauthorized access, use, and disclosure of data, and provisions for enforcement and redress. Buchanan (2011) proposes three ethical principles, namely "respect for persons, beneficence, and justice" (p. 84). Not only should individuals be regarded as autonomous agents, but vulnerable individuals with diminished or constrained autonomy (including students) should be protected from harm and risk.

Although many of the above recommendations provide useful pointers for learning analytics, seeing learning analytics as a moral practice with its primary contribution of increasing the effectiveness of learning necessitates a different (but not contradictory) set of principles.

Principles for an Ethical Framework for Learning Analytics

Our approach holds that an institution's use of learning analytics is going to be based on its *understanding* of the scope, role, and boundaries of learning analytics and a set of moral beliefs founded on the respective regulatory and legal, cultural, geopolitical, and socioeconomic contexts. Any set of guidelines concerned with the ethical dilemmas and challenges in learning analytics will necessarily also be based on a set of epistemological assumptions. As such, it would be almost impossible to develop a set of universally valid guidelines that could be equally applicable within any context. However, it should be possible to develop a set of general principles from which institutions can develop their own sets of guidelines depending on their contexts.

We propose here a number of principles as a guiding framework for considering learning analytics as moral practice.

Principle 1: Learning Analytics as Moral Practice

In response to the increasingly analytic possibilities facing the current institution of higher education, learning analytics should do much more than contribute to a "data-driven university" or lead to a world where we are "living under the sword of data." We agree with Biesta (2007) that

> evidence-based education seems to favor a technocratic model in which it is assumed that the only relevant research questions are about the effectiveness of educational means and techniques, forgetting, among other things, that what counts as "effective" crucially depends on judgments about what is educationally desirable. (p. 5)

Education cannot and should not be understood as "as an intervention or treatment because of the noncausal and normative nature of educational practice and because of the fact that the means and ends in education are internally related" (Biesta, 2007, p. 20). Learning analytics should not only focus on what is effective, but also aim to provide relevant pointers to decide what is appropriate and morally necessary. Education is primarily a moral practice, not a causal one. Therefore, learning analytics should function primarily as a *moral* practice resulting in understanding rather than measuring (Reeves, 2011).

Principle 2: Students as Agents

In stark contrast to seeing students as producers and sources of data, learning analytics should engage students as collaborators and not as mere recipients of interventions and services (Buchanan, 2011; Kruse & Pongsajapan, 2012). Not only should students provide informed consent regarding the collection, use, and storage of data, but they should also voluntarily collaborate in providing data and access to data to allow learning analytics to serve *their* learning and development, and not just the efficiency of institutional profiling and interventions (also see Subotzky & Prinsloo, 2011). Kruse and Pongsajapan (2012) propose a "student-centric," as opposed to an "intervention-centric," approach to learning analytics. This suggests the student should be seen

> as a co-interpreter of his own data—and perhaps even as a participant in the identification and gathering of that data. In this scenario, the student becomes aware of his own actions in the system and uses that data to reflect on and potentially alter his behavior. (pp. 4-5)

Valuing students as agents, making choices and collaborating with the institution in constructing their identities (however transient), can furthermore be a useful (and powerful) antidote to the commercialization of higher education (see, e.g., Giroux, 2003) in the context of the impact of skewed power relations, monitoring, and surveillance (Albrechtslund, 2008; Knox, 2010a, 2010b).

Principle 3: Student Identity and Performance Are Temporal Dynamic Constructs

Integral in learning analytics is the notion of student identity. It is crucial to see student identity as a combination of permanent and dynamic attributes. During students' enrollment, their identities are in continuous flux, and as such they find themselves in a "Third Space" where their identities and competencies are in a permanent liminal state (Prinsloo, Slade, & Galpin, 2012). The ethical implications of this are that learning analytics provides a snapshot view of a learner at a particular time and context. This not only necessitates the need for longitudinal data (Reeves, 2011) but also has implications for the storage and permanency of data. Mayer-Schönberger (2009, p. 12) warns that forgetting is a "fundamental human capacity." Students should be allowed to evolve and adjust and learn from past experiences without those experiences, because of their digital nature, becoming permanent blemishes on their development history. Student profiles should not become "etched like a tattoo into . . . [their] digital skins" (Mayer-Schönberger, 2009, p. 14). Data collected through learning analytics should therefore have an agreed-on life span and expiry date, as well as mechanisms for students to request data deletion under agreed-on criteria.

Principle 4: Student Success Is a Complex and Multidimensional Phenomenon

Although one of the benefits of learning analytics is to contribute to a better understanding of student demographics and behaviors (Bichsel, 2012), it is important to see student success is the result of "mostly non-linear, multidimensional, interdependent interactions at different phases in the nexus between student, institution and broader societal factors" (Prinsloo, 2012). Although learning analytics offer huge opportunities to gain a more comprehensive understanding of student learning, our data are incomplete (e.g., Booth, 2012; Mayer-Schönberger, 2009; Richardson, 2012a, 2012b) and "dirty" (Whitmer et al., 2012) and our analyses vulnerable to misinterpretation and bias (Bienkowski et al., 2012; Campbell et al., 2007; May, 2011).

Principle 5: Transparency

Important for learning analytics as moral practice is that higher education institutions should be transparent regarding the purposes for which data will be used and under which conditions, who will have access to data, and the measures through which individuals' identity will be protected. The assumption that participating in public online forums provides blanket permission for use of data should not be acceptable (Buchanan, 2011). Higher education institutions have an obligation to protect participant data on the institutional LMS and also to inform students of possible risks when teaching and learning occur outside the boundaries of institutional jurisdiction.

Principle 6: Higher Education Cannot Afford to Not *Use Data*

The previous five principles provide guidance for those higher education institutions using or planning to use data. The sixth principle makes it clear that higher education institutions cannot afford to not use learning analytics. The triggers for adopting learning analytics will depend on an institution's answer to the earlier question regarding its main purpose. Whether their purpose is to earn profit or to improve outcomes for students, institutions should use available data to better understand and then engage with, and indeed ameliorate, the outcomes (Bienkowski et al., 2012; Campbell et al., 2007). Ignoring information that might actively help to pursue an institution's goals seems shortsighted to the extreme. Institutions are accountable, whether it is to shareholders, to governments, or to students themselves. Learning analytics allows higher education institutions to assist all stakeholders to penetrate "the fog that has settled over much of higher education" (Long & Siemens, 2011, p. 40).

Considerations for Learning Analytics as Moral Practice

In line with the proposal by Ess et al. (2012), we propose a number of considerations rather than a code of practice to allow for flexibility and the range of contexts in which they might need to be applied. The following considerations are structured to address issues regarding benefits, consent, vulnerability and harm, data, and governance and resource allocation.

Who Benefits and Under What Conditions?

The answer to this question is *the* basis for considering the ethical dimensions of learning analytics. From the literature review, it is clear that *both* students and the institution should benefit, and that the most benefit is derived when students and institutions collaborate as stakeholders in learning analytics. Students are not simply recipients of services or customers paying for an education. They are and should be active agents in determining the scope and purpose of data harvested from them and under what conditions (e.g., deidentification). On the other hand, it is clear that to deliver increasingly effective and appropriate learning and student support, higher education institutions need to optimize the selection of data harvested and analyzed. We strongly suspect that students should be informed that, to deliver a personalized and appropriate learning experience, higher education needs not only to harvest data but also to ensure that deidentification of data should not hamper personalization. Agreeing on the need for and purpose of harvesting data under certain provisions provides a basis of trust between the institution and students. Both parties to the agreement realize that the veracity and comprehensiveness of data will allow optimum personalization, appropriate learning and support, and cost-effectiveness.

Conditions for Consent, Deidentification, and Opting Out

Although informed consent is an established practice in research, the same cannot be said for use of student data within other educational contexts. Given that the scope and nature of available data have changed dramatically, we should revisit the notion of informed consent in the field of learning analytics.

Is informed consent a sine qua non, or are there circumstances in which other principles override the need for informed consent? There are many examples in different fields (e.g., bioethics) where the principle of informed consent can be waived under predetermined circumstances or if existing legislation is sufficient (e.g., data protection legislation). In extreme cases, informed consent may be forgone if the benefit to the many exceeds the needs of the individual. In the context of learning analytics, we might suggest that there are few, if any, reasons not to provide students with information regarding the uses to which their data might be put, as well as the models used (as far as they may be known at that time), and to establish a system of informed consent. Given the continuing advances in technology and our understanding of the effective applications of learning analytics, this consent may need to be refreshed on a regular basis. As Herring (2002) states, the changing composition of student groupings over time suggests that obtaining informed consent may be problematic. She suggests the need to achieve a reasonable balance between allowing quality research to be conducted and protecting users from potential harm. In practice, this may translate to provision of a broad definition of the range of potential uses to which a student's data may be put, some of which may be less relevant to the individual.

Buchanan (2011), referring to the work of Lawson (2004), suggests a nuanced approach to consent that offers students a range of options for withholding (partial) identification of individuals where they are part of a published study. In light of this, it seems reasonable to distinguish between analyzing and using anonymized data for reporting purposes to regulatory bodies or funding purposes and other work on specific aspects of student engagement. In the context of reporting purposes, we support the notion that the benefit for the majority supersedes the right of the individual to withhold permission for use of his or her data. Students may, however, choose to opt out of institutional initiatives to personalize learning—on the condition that students are informed and aware of the consequences of their decision.

Institutions should also provide guarantees that student data will be permanently deidentified after a certain number of years, depending on national legislation and regulatory frameworks.

Vulnerability and Harm

Definitions. How do we define vulnerability and harm and prevent potential harm, not only to students but to all stakeholders? In considering this issue, we might think about aspects such as implicit or explicit discrimination (whereby one student receives, or does not receive, support based on what might externally be considered to be a random personal characteristic), labeling (where students may be branded according to some combination of characteristics and treated—potentially for the duration of their studies—as somehow different from others), and the validity of treating groupings of students in a particular way based on assumptions made about their shared characteristics. One way to address the potential role of bias and stereotyping is to adopt a position of "Rawlsian blindness" (Harvey, 2012, para. 12) where students' demographic and prior educational records are not used from the outset to predict their chances of success and failure. On the other hand, is it ethical to ignore the predictive value of research evidence in particular contexts?

We suggest that the potential for bias and stereotyping in predictive analysis should be foregrounded in institutional attempts to categorize students' risk profiles. Institutions should provide additional opportunities for these students either to prove the initial predictive analyses wrong or incomplete or to redeem themselves despite any initial institutional doubt regarding their potential. In determining what might constitute vulnerability in the context of learning analytics, institutions should aim to ensure that analyses are conducted on robust and suitably representative data sets.

Redress for students. If a system of transparency and informed consent is adopted, it might be argued that the potential for allegations of misuse and harm is minimized. It is unlikely though that approaches can be fully comprehensive in their consideration of potential (future) scenarios, and it is feasible that students may argue that they have been disadvantaged (perhaps by not receiving the perceived advantages that other students have). Bollier (2010) provides an example of a future whereby low-risk (and therefore low-cost) students may seek preferential treatment (reduced entrance requirements, perhaps) at a "cost" to perceived high-risk students. At this stage, it is difficult to assess longer-term implications, although most higher education institutions will have in place clear complaints and appeals systems, which will perhaps warrant revision.

Redress for institutions. Conversely, if systems and approaches are transparent, there is increased potential for student abuse of the system. What recourse do institutions have when students provide false or incomplete information that may provide them with additional support at a cost to the institution (and to other students)? Student regulations typically contain statements that allow the institution to terminate registration if information given is untrue or misleading, and there are other less draconian measures that might also be adopted.

Data Collection, Analyses, Access, and Storage

Collection of data. In collecting data from disparate sources, institutions need to take due care to ensure not to "amplify error" resulting from the "different standards, owners and levels of access" on different sites (Ferguson, 2012, p. 6). It is generally accepted that data on the institutional LMS provide an incomplete picture of learners' learning journeys. As learners' digital networks increasingly include sources outside of the LMS, institutions may utilize data from *outside* the LMS (e.g., Twitter and Facebook accounts, whether study related or personal) to get more comprehensive pictures of students' learning trajectories. The inclusion of data from sites not under the jurisdiction of an institution raises a number of concerns given that universities have no control of external sites' policies, and the authentication of student identity is more problematic. Students have the right to be informed on the sites used to gather information and to give informed consent regarding the scope and right of the institution to harvest, analyze, and use data from such sources. Registration information should be explicit regarding the broader uses to which student data may be put.

Analyses of data. Institutions should commit themselves to take due care to prevent bias and stereotyping, always acknowledging the incomplete and dynamic nature of individual identity and experiences. Algorithms used in learning analytics inherently may reflect and perpetuate the biases and prejudices in cultural, geopolitical, economic, and societal realities. As Subotzky and Prinsloo (2011, p. 182) state, our predictive models explain only "a portion of the wide range of behaviours that constitute the universe of social interactions" between students and institution. Students and the institution therefore share a mutual responsibility, which "depends upon mutual engagement, which, in turn, depends on actionable mutual knowledge" (Subotzky & Prinsloo, 2011, p. 183).

Care should be taken to ensure that the contexts of data analyzed are carefully considered before tentative predictions and analyses are made. Data harvested in one context may not be directly transferable to another. Predictive models and algorithms should take reasonable care to prevent "autopropaganda" (Pariser, 2011) and allow for serendipity and for students to act outside of modeled profiles and suggestions for action.

Access to data. In line with Kruse and Pongsajapan's (2012) proposal for "student-centric" learning analytics, we propose that students have a right to be assured that their data will be protected against unauthorized access and that their informed consent (as discussed above) is guaranteed when their data are used. Given the unequal power relationship between student and institution, the institution should take steps to safeguard access to student data and provide students with processes for redress should unauthorized persons gain access to their personal data.

In practice, owing to the nature of regulatory, funding, and accreditation frameworks, a variety of stakeholders do access student data. In cases where employers fund students' study, sponsors may have rights to data relating to student progress. It is suggested that students have ready access to their personalized stored data, as well as an overview of those stakeholders granted access to specific data sets. Institutions should also take reasonable steps to make students aware of the scope and nature of their data trails when using a range of social networking sites in the course of their studies. Although institutions cannot be held responsible for the level of students' digital literacy, there is possibly a strong case to support digital literacy as an integral part of the attributes of graduates (Nawaz & Kundi, 2010). When teaching and learning opportunities incorporate social networks outside of the institutional LMS, institutions should also ensure that learners are explicitly informed of the public nature and possible misuse of information posted on these sites, and instructors should consider the ramifications before using such sites.

Preservation and storage of data. Although ownership of data in non-Internet-based research is fairly straightforward, ownership of data obtained from the Internet is less clear, and the lack of international boundaries "confound ownership, as databanks, datasets, servers, find their homes across borders" (Buchanan, 2011, p. 98). Institutions should provide guarantees and guidelines with regard to the preservation and storage of data in line with national and international regulatory and legislative frameworks. Students should be informed that their data will be encrypted. Many countries have put in place legal frameworks to ensure that individuals can apply for the correction or permanent deletion of any personal information held about them that may be inaccurate, misleading, or outdated. For example, in the South African context, the Promotion of Access to Information Act 2 of 2000 provides individuals with "a level of direct influence over the processing of their personal information" (KPMG, 2010). Petersen (2012) suggests,

> One of the first tasks of a data-governance body is to inventory campus data sources and create a data-classification system that sorts data into categories that identify the necessary level of protection. This process may necessarily be different for a private versus a public institution, since state laws or regulations may require that certain information be available to the public. A typical classification scheme for a public college or university might include the categories of (1) public, (2) non-public, and (3) sensitive or regulated. (P. 46)

Governance and Resource Allocation

Higher education institutions should, depending on national legislative and regulatory frameworks, ensure the effective governance and stewardship of data. Petersen (2012, P. 46) states, "The most important step that any campus can take is to create a comprehensive data-governance structure to address all the types of data used in various situations." This implies not only the strategic conceptualization of learning analytics but also appropriate structural and resource allocation. The allocation of resources inevitably relates to purpose and benefit. Before embarking on the application of a learning analytics approach, the institution (or faculty) should be clear about what its key drivers for success are, what constraints exist, and which conditions must be met.

Conclusions

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

Reflecting on the future directions for research ethics in networked environments ("research ethics 2.0"), Buchanan (2011) remarks, "As social networking, hyper-blogging, folksonomies, Wikis, etc., continue to change social interaction, research itself and this research ethics *must* change." Researchers and ethics boards should "work in tandem to forge the next generation of research ethics, one that still embraces core principles while creating new opportunities for important research endeavors" (p. 103).

Learning analytics is primarily a moral and educational practice, serving better and more successful learning. The inherent peril and promise of having access to and analyzing "big data" (Bollier, 2010) necessitate a careful consideration of the ethical dimensions and challenges of learning analytics. The proposed principles and considerations included within this article provide an ethical framework for higher education institutions to offer context-appropriate solutions and strategies to increase the quality and effectiveness of teaching and learning.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Adams, C., & Van Manen, M. (2006). Embodiment, virtual space, temporality and interpersonal relations in online writing. *College Quarterly*, *9*(4), 1-18. Retrieved from http://www.maxvanmanen.com/files/2011/04/2007-Embodiment-Virtual-Space.pdf

Albrechtslund, A. (2008). Online social networking as participatory surveillance. *First Monday*, *13*(3). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2142/1949

Andrejevic, M. (2011). Surveillance and alienation in the online economy. *Surveillance & Society*, *8*(3), 278-287.

Apple, M. W. (2004). *Ideology and curriculum* (3rd ed.). New York, NY: Routledge Falmer.

Arrington, M. (2010, 10 January). *Facebook CEO Mark Zuckerberg: TechCrunch interview at the Crunchies*. Retrieved from http://www.youtube.com/watch?v=LoWKGBloMsU

Bichsel, J. (2012, August). *Analytics in higher education: Benefits, barriers, progress, and recommendations (Research report)*. Louisville, CO: EDUCAUSE Center for Applied Research. Available from http://net.educause.edu/ir/library/pdf/ERS1207/ers1207.pdf

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Draft paper submitted to the Office of Educational Technology, U.S. Department of Education. Retrieved from http://evidenceframework.org/wp-content/uploads/2012/04/EDM-LA-Brief-Draft_4_10_12c.pdf

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*(1), 1-22. Retrieved from http://www.vbsinternational.eu/files/media/research_article/Evidencebased_education_Biesta1.pdf

Bollier, D. (2010). *The promise and peril of big data*. Washington, DC: Aspen Institute. Retrieved from http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

Booth, M. (2012, July/August). Learning analytics: The new black. *EDUCAUSE Review*, *47*(4), 52-53.

Brah, A. (1996). *Cartographies of diaspora: Contesting identities*. London, UK: Routledge.

Buchanan, E. A. (2011). Internet research ethics: Past, present and future. In M. Consalvo & C. Ess (Eds.), *The handbook of Internet studies* (pp. 83-108). Oxford, UK: John Wiley.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review*, *42*(4), 40-57. Retrieved from http://net.educause.edu/ir/library/pdf/ERM0742.pdf

Cokins, G. (2009). *Performance management: Integrating strategy execution, methodologies, risk, and analytics*. Hoboken, NJ: John Wiley.

Coll, S., Glassey, O., & Balleys, C. (2011). Building social networks ethics beyond "privacy": A sociological perspective. *International Review of Information Ethics*, *16*(12), 48-53.

Cooper, N. (2009). Workforce demographic analytics yield health-care savings. *Employment Relations Today*, *36*(3), 13-18.

Davenport, T. H., Harris, J., & Shapiro, J. (2010, October). Competing on talent analytics. *Harvard Business Review*, pp. 2-6. Retrieved from https://www.growthresourcesinc.com/pub/res/pdf/HBR_Competing-on-Talent-Analytics_Oct-2010.pdf

Dawson, S. (2006). The impact of institutional surveillance technologies on student behaviour. *Surveillance & Society*, *4*(1/2), 69-84.

Dillon, M., & Loboguerrero, L. (2008). Biopolitics of security in the 21st century: An introduction. *Review of International Studies*, *34*, 265-292. doi:10.1017/S0260210508008024

Dolan, J. W. (2008). Involving patients in decisions regarding preventive health interventions using the analytic hierarchy process. *Health Expectations*, *3*, 37-45.

Epling, M., Timmons, S., & Wharrad, H. (2003). An educational Panopticon? New technology, nurse education and surveillance. *Nurse Education Today*, *23*(6), 412-418. doi:org/10.1016/S0260-6917(03)00002-9

Ess, C., Buchanan, E., & Markham, A. (2012). *Ethical decision-making and Internet research: 2012 recommendations from the AOIR Ethics Working Committee*. Unpublished manuscript.

Ferguson, R. (2012, March). *The state of learning analytics in 2012: A review and future challenges (Tech. Rep. No. KMI-12-01)*. Milton Keynes, UK: The Open University, Knowledge Media Institute. Retrieved from http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Giroux, H. A. (2003). Selling out higher education. *Policy Futures in Education*, *1*(1), 179-200.

Hall, R., & Stahl, B. (2012). Against commodification: The university, cognitive capitalism and emerging technologies. *TripleC*, *10*(2), 184-202.

Harvey, A. (2012, May 16). Student entry checks by universities carry risks. *The Australian*. Retrieved from http://www.theaustralian.com.au/higher-education/opinion/student-entry-checks-by-universities-carry-risks/story-e6frgcko-1226356721317

Herring, S. (2002). Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology*, *36*(1), 109-168.

Johnson, L., Adams, S., & Cummins, M. (2012). *The NMC horizon report: 2012 higher education edition*. Austin, TX: New Media Consortium.

Knox, D. (2010a). A good horse runs at the shadow of the whip: Surveillance and organizational trust in online learning environments. *Canadian Journal of Media Studies*. Retrieved from http://cjms.fims.uwo.ca/issues/07-01/dKnoxAGoodHorseFinal.pdf

Knox, D. (2010b, October). *Spies in the house of learning: A typology of surveillance in online learning environments*. Paper presented at Edge2010, St. Johns, Canada. Retrieved from http://www.mun.ca/edge2010/wp-content/uploads/Knox-Dan-Spies-In-the-House.pdf

KPMG. (2010). *The protection of personal information bill—Principle 8: Data subject participation*. Retrieved from http://www.kpmg.com/za/en/issuesandinsights/articlespublications/protection-of-personal-information-bill/pages/principle-8%E2%80%93data-subject-participation.aspx

Kruse, A., & Pongsajapan, R. (2012). *Student-centered learning analytics*. Retrieved from https://cndls.georgetown.edu/m/documents/thoughtpaper-krusepongsajapan.pdf

Land, R., & Bayne, S. (2005). Screen or monitor? Surveillance and disciplinary power in online learning environments. In R. Land & S. Bayne (Eds.), *Education in cyberspace* (pp. 165-178). London, UK: Routledge Falmer.

Lawson, D. (2004). Blurring the boundaries: Ethical considerations for online research using synchronous CMC forums. In E. A. Buchanan (Ed.), *Readings in Virtual Research Ethics: Issues and Controversies* (pp. 80-100). Hershey, PA and London, UK: Information Science Publishing.

Livingstone, S. (2005). In defence of privacy: Mediating the public/private boundary at home. In S. Livingstone (Ed.), *Changing Media–Changing Europe Series: Vol. 2. Audiences and publics: When cultural engagement matters for the public sphere* (pp. 163-184). Bristol, UK: Intellect Books.

Lombardo, P. A., & Dorr, G. M. (2006). Eugenics, medical education, and the Public Health Service: Another perspective on the Tuskegee syphilis experiment. *Bulletin of the History of Medicine*, *80*, 291-316.

Long, P., & Siemens, G. (2011, September/October). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 31-40.

Lyon, D. (2007). *Surveillance studies: An overview*. Cambridge, UK: Polity Press.

May, H. (2011). *Is all the fuss about learning analytics just hype?* Retrieved from http://www.loomlearning.com/2011/analytics-schmanalytics

Mayer-Schönberger, V. (2009). *Delete: The virtue of forgetting in the digital age*. Princeton, NJ: Princeton University Press.

Nawaz, A., & Kundi, G. M. (2010). Digital literacy: An analysis of the contemporary paradigms. *Journal of Science and Technology Education Research*, *1*(2), 19-29. Retrieved from http://www.academicjournals.org/IJSTER/PDF/Pdf2010/July/Nawaz%20and%20Kundi.pdf

Nelson, K., & Creagh, T. (2012). *The good practice guide for safeguarding student learning engagement. Unpublished draft report*, Queensland University of Technology, Queensland, Australia.

Oblinger, D. G. (2012, July/August). Let's talk analytics. *EDUCAUSE Review*, *47*(4), 10-13.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London, UK: Viking.

Parry, M. (2012, July 18). Big data on campus. *New York Times*. Retrieved from http://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html?_r=1&pagewanted=all

Petersen, R. J. (2012, July/August). Policy dimensions of analytics in higher education. *EDUCAUSE Review*, *47*(4), 44-49.

Prinsloo, P. (2012, May). *ODL research at Unisa*. Presentation at the School of Management Sciences, University of South Africa, Pretoria.

Prinsloo, P., Slade, S., & Galpin, F. A. V. (2012, 29 April-2 May). *Learning analytics: Challenges, paradoxes and opportunities for mega open distance learning institutions*. Paper presented at the 2nd International Conference on Learning Analytics and Knowledge (LAK12), Vancouver, Canada. doi:0.1145/2330601.2330635

Reeves, T. C. (2011). Can educational research be both rigorous and relevant? *Educational Designer*, *1*(4), 1-24. Retrieved from http://www.educationaldesigner.org/ed/volume1/issue4/article13/pdf/ed_1_4_reeves_11.pdf

Richardson, W. (2012a, July 14). *Valuing the immeasurable [Blog post]*. Retrieved from http://willrichardson.com/post/27223512371/valuing-the-immeasurable

Richardson, W. (2012b, August 3). *The "immeasurable" part 2 [Blog post]*. Retrieved from http://willrichardson.com/post/28626310240/the-immeasurable-part-2

Seifert, J. W. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. *Government Information Quarterly*, *21*, 461-480. doi:10.1016/j.giq.2004.08.006

Sen, A. (2006). *Identity and violence: The illusion of destiny*. Harmondsworth, UK: Penguin.

Siemens, G. (2012). *Learning analytics: Envisioning a research discipline and a domain of practice*. Paper presented at the 2nd International Conference on Learning Analytics and Knowledge (LAK12), Vancouver, Canada. Retrieved from http://learninganalytics.net/LAK_12_keynote_Siemens.pdf

Snaedal, J. (2002). The ethics of health sector databases. *eHealth International*, 1, 6. doi:10.1186/1476-3591-1-6. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC135527/

Subotzky, S., & Prinsloo, P. (2011). Turning the tide: A socio-critical model and framework for improving student success in open distance learning at the University of South Africa. *Distance Education*, *32*(2), 177-193.

Varvel, V. E., Montague, R.-A., & Estabrook, L. S. (2007). Policy and e-learning. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage handbook of e-learning research* (pp. 269-285). London, UK: Sage.

Wagner, E., & Ice, P. (2012, July/August). Data changes everything: Delivering on the promise of learning analytics in higher education. *EDUCAUSE Review*, *47*(4), 33-42.

Wel, L. V., & Royakkers, L. (2004). Ethical issues in web data mining. *Ethics and Information Technology*, *6*(2), 129-140. doi:10.1023/B:ETIN.0000047476.05912.3d

Whitmer, J., Fernandes, K., & Allen, W. R. (2012, August 13). Analytics in progress: Technology use, student characteristics, and student achievement. *EDUCAUSE Review*. Retrieved from http://www.educause.edu/ero/article/analytics-progress-technology-use-student-characteristics-and-student-achievement

Author Biographies

Sharon Slade is a senior lecturer at the Open University in the UK with research interests in online teaching and learning and the use of learning analytics to support student retention.

Paul Prinsloo is an education consultant and researcher at the University of South Africa. His passion and research interests include curriculum development, student success and retention, learning analytics and learning design. He blogs at http://opendistanceteachingandlearning.wordpress.com.