

Ethical Considerations in the Practical Application of the Unisa Socio-Critical Model of Student Success

Angelo Fynn

Volume 17, Number 6, December 2016

URI: <https://id.erudit.org/iderudit/1064647ar>

DOI: <https://doi.org/10.19173/irrodl.v17i6.2812>

[See table of contents](#)

Publisher(s)

Athabasca University Press (AU Press)

ISSN

1492-3831 (digital)

[Explore this journal](#)

Cite this article

Fynn, A. (2016). Ethical Considerations in the Practical Application of the Unisa Socio-Critical Model of Student Success. *International Review of Research in Open and Distributed Learning*, 17(6), 206–220.
<https://doi.org/10.19173/irrodl.v17i6.2812>

Article abstract

The prediction and classification of student performance has always been a central concern within higher education institutions. It is therefore natural for higher education institutions to harvest and analyse student data to inform decisions on education provision in resource constrained South African environments. One of the drivers for the use of learning and academic analytics is the pressures for greater accountability in the areas of improved learning outcomes and student success. Added to this is the pressure on local government to produce well-educated populace to participate in the economy. The counter discourse in the field of big data cautions against algocracy where algorithms takeover the human process of democratic decision making. Proponents of this view argue that we run the risk of creating institutions that are beholden to algorithms predicting student success but are unsure of how they work and are too afraid to ignore their guidance. This paper aims to discuss the ethics, values, and moral arguments revolving the use of big data using a practical demonstration of learning analytics applied at Unisa.

Copyright © Angelo Fynn, 2016



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

December – 2016

Ethical Considerations in the Practical Application of the Unisa Socio-Critical Model of Student Success

Angelo Fynn
University of South Africa

Abstract

The prediction and classification of student performance has always been a central concern within higher education institutions. It is therefore natural for higher education institutions to harvest and analyse student data to inform decisions on education provision in resource constrained South African environments. One of the drivers for the use of learning and academic analytics is the pressures for greater accountability in the areas of improved learning outcomes and student success. Added to this is the pressure on local government to produce well-educated populace to participate in the economy. The counter discourse in the field of big data cautions against algocracy where algorithms takeover the human process of democratic decision making. Proponents of this view argue that we run the risk of creating institutions that are beholden to algorithms predicting student success but are unsure of how they work and are too afraid to ignore their guidance. This paper aims to discuss the ethics, values, and moral arguments revolving the use of big data using a practical demonstration of learning analytics applied at Unisa.

Keywords: learning analytics, academic analytics, algocracy, higher education

Introduction

This paper aims to conduct a critical discussion of the broader ethical and practical considerations to be considered in the implementing the Socio-critical model (2011). The paper outlines the basic premise of the Socio-critical model and highlights key arguments raised by authors in the field of ethics in big data mining.

Central to the implementation of the Socio-critical model of student success is the use of effective, contextualized learning analytics. Prinsloo and Slade (2014a) point out that the higher education system has always made use of historical student data to plan institutional responses to student learning and support needs (p.197). Technological progression has allowed for a shift toward more dynamic datasets that provide real time, more complete student profiles to facilitate tailored support and customized learning pathways (Prinsloo & Slade, 2014a; Subotzky & Prinsloo, 2011; Venter & Prinsloo, 2011).

There are two overarching concerns raised when the ethics of the use of analytics in Higher Education, namely, privacy and the threat of algocracy (Danaher, 2014; Morozov, 2013). These authors caution against the increasing reliance on the use of big data without the necessary interrogation of the underlying assumptions and institutional practices (Danaher, 2014; Morozov, 2013). In particular, these authors call for interrogation of practices that inform the method of data collection, the ethics of data ownership and the application of the data outcomes to regulate behaviour.

On Big Data: Academic and Learning Analytics

The use of learning analytics in education is relatively new and the exact definition of the term is still being developed. Analytics, generally speaking, is a data-driven approach to decision making based on the discovery and communication of meaningful patterns of data in the business or educational context (Jayaprakash, Moody, Eitel, Regan, & Baron, 2014). Literature on analytics in education includes references to *academic analytics*, *predictive analytics*, *educational data mining*, *social learning analytics*, and *learning analytics* among others (Jayaprakash et al., 2014). While each of the applications of analytics bears discussion, the focus of this paper is on learning analytics.

Learning analytics, if defined broadly, focuses on learner produced-data and analysis models for predicting individual learning (Siemens, 2010). The goals of learning analytics can be divided into four broad outcome areas, namely, prediction, personalisation, intervention, and information visualisation. The predictive function of learning analytics strives to identify at risk students in terms of dropout and failure while personalisation focuses on providing students with tailored learning pathways and assessments.

Although not proposed as the panacea for student risk and throughput, the emergence of big data and analytic technology provide new tools to address the issues at hand (Jayaprakash et al., 2014). Using massive datasets drawn from student systems representative of interactions between student and institutional systems it is now possible to accurately predict whether a student is likely to fail or dropout of a course or degree program (Jayaprakash et al., 2014).

The use of analytics to predict student success often operates on a continuum of responsible resource usage and the concept of educational triage (Prinsloo & Slade, 2014b; Simpson, 2006). The former is typically aimed at reducing resource waste on students who have a low chance of succeeding while the

latter is concerned with providing key support to students who most needed and stand to benefit from the intervention.

One of the more well-known applications of learning analytics is the Course Signals (CS) project at Purdue (Arnold, Hall, Street, Lafayette, & Pistilli, 2012). The CS project aimed to facilitate student integration into the institution by allowing faculty to send personalized emails to students that contain information about current performance in a given course. The project also allowed faculty to encourage students to visit various resources on the assumption that student engagement with support services can mitigate risk. Finally, it employed learning analytics to allow for the real time integration of data on student performance (Arnold et al., 2012). The stated objectives of the CS project were premised on the work of Tinto (as cited in Arnold et al, 2012) which operates on the premise that the welfare of the student should supersede that of the institution. Tinto (as cited in Arnold et al, 2012)) further argues that interventions should be focused on all students at an institution and that the interventions aimed at enhancing student success should be focused on assisting students to integrate academically into an institution (Arnold et al., 2012). The position that students should integrated into the institution is in contrast to the position taken by Subotzky and Prinsloo (2011) which holds that success requires adaptation by both student and institution.

On The Notion of Student Success at Unisa

Subotzky and Prinsloo (2011) have proposed a broad definition of success premised on their Socio-critical model for student success, described later in this paper. Subotzky and Prinsloo (2011) define student success as “course success leading to graduation and time to completion within the expected minimum time appropriate to the qualification types within the ODL context” (p.188).

Subotzky and Prinsloo (2011) go on to elaborate on this definition by adding that a positive student experience and high levels of satisfaction throughout the student walk at the institution is a key factor in ascertaining institutional success (Subotzky & Prinsloo, 2011). A third aspect of student success relates to the “successful fit between students’ graduate attributes and the requirements of the workplace, civil society and democratic, participative citizenship” (p.188). The definition by Subotzky and Prinsloo (2011) also argues for a definition of success outside of the traditional graduate success model by accommodating students who do not fit the traditional profile. They propose that students who are pursuing learning for its intrinsic value or who shift to complete studies at other institutions should be integrated into the definition of success. The rationale behind the argument is that it is possible that they have derived benefit and enrichment from their exposure to the ODL context.

Subotzky and Prinsloo (2011) highlight the continuing impact of the legislative, educational, and epistemological impact of the colonial and apartheid frameworks on South African higher education. They argue that any analysis of educational performance, of both individual learners and institutions, must take a socio-critical stance to fully understand the concept of success (Subotzky & Prinsloo, 2011). Furthermore, they highlight that Unisa’s student profile is predominantly non-traditional, older, or part-time students (Subotzky & Prinsloo, 2011). This implies that we cannot import our understanding of success from other institutions within South Africa (Subotzky & Prinsloo, 2011). In response to the social, historic, political, economic, and cultural challenges facing South African higher education, Subotzky and Prinsloo (2011) propose a definition of student success. The proposed model consists of five constructs, namely, the situated agents (student and institution); the interaction between the student, institution and networks (the student walk); capital; habitus; and the domains and modalities of transformation (see Figure 1).

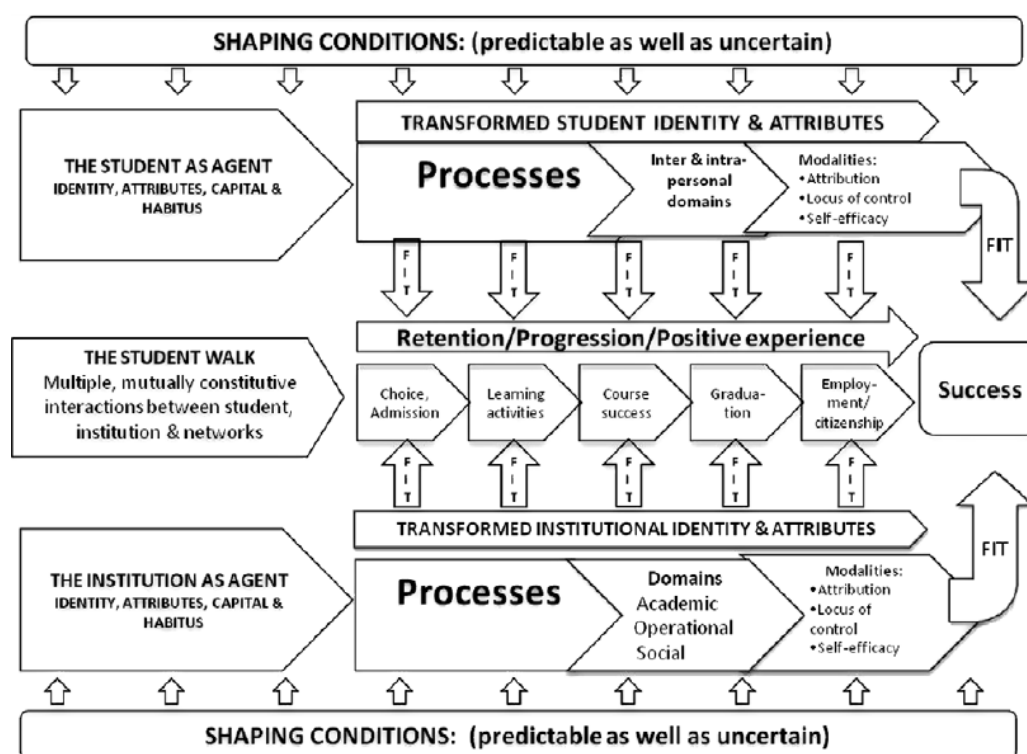


Figure 1. The Socio-critical model for student success.

The Socio-critical model described above was integrated into a student risk tracking system that is used to provide near real time analytics of students propensity for failure and dropout. The primary purpose of the project is the development of a near real time, analytics system that is fully integrated into the information management systems of the institution. A secondary outcome of the project is the development of strategies for integrating the various information stores at Unisa into a coherent whole that will facilitate intelligent decision making.

In addition to the theoretical model for student success, Subotzky and Prinsloo (2011) proposed a simplified model for implementing the collection and analysis of data to test and monitor the model. Figure 2 below visually outlines the primary elements of the process for implementing the student success proposed by Subotzky and Prinsloo (2011). The process shown below provides a broad strategy for the development and implementation of the Socio-critical model of success proposed above.

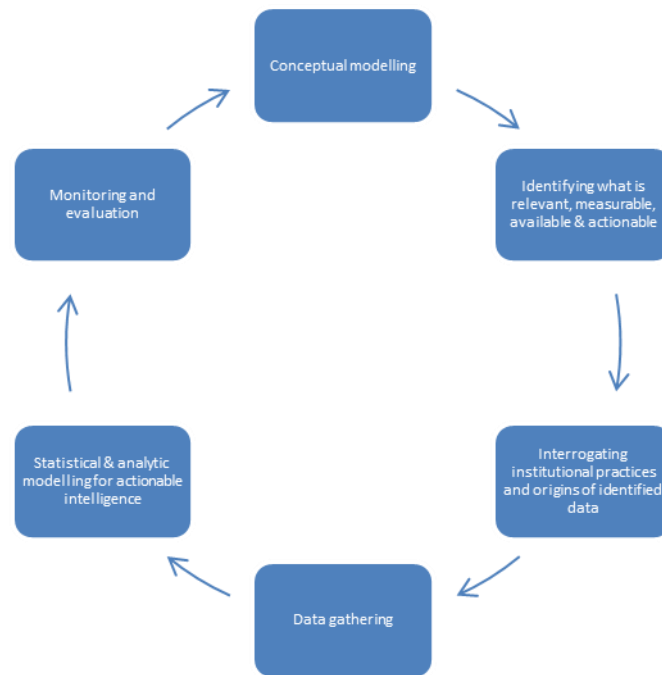


Figure 2. The main elements and process of managing student success at Unisa (Subotzky & Prinsloo, 2011).

The modelling process set out by Subotzky and Prinsloo (2011) provides a useful, if broad, overview of the process to follow when developing a process focussed on developing a data-based model of student success. The model, however, requires adjustment in light of the concerns raised by Morozov (2013) and others. Specifically, the cyclical nature of the modelling process requires greater emphasis and the addition of a consultation process following the identification of what is relevant, measurable, available and actionable (Subotzky & Prinsloo, 2011). The additional process would focus on a consultative interrogation of the underlying assumptions and institutional practices that inform the method of data collection. Additionally the consultative process would have to address the ethics of data ownership and the application of the data outcomes to regulate behavior as proposed by Danaher (2014) and Morozov (2013).

Technical Limitations on Applying Analytics at Unisa

The current Unisa system harvests from five distinct areas, namely, admission, learning activity, course success, graduation and employment to model student success (Prinsloo et al., 2015). The learning activities are naturally geared toward learning analytics while the admissions, course success, graduation, and employment are in line with the aims of academic analytics (Prinsloo et al., 2015). The data is harvested from five distinct data warehouses that are structured along the major functions of the institution:

1. student administration,
2. learner management,
3. human resources,
4. finance, and

5. space and the built environment (Prinsloo et al., 2015).

The disparate nature of the data sources requires specialized expertise to translate the business rules embedded in the design and function of each data warehouse into extracts that can be utilised within an analytic framework (Prinsloo et al., 2015). The inherent risk with this model is that with each new configuration of the data, the original context, assumptions and ideologies underpinning the data are lost (Prinsloo et al., 2015). The disparate nature of the data sources, combined with the risks inherent in reconfiguring data further emphasises the need for interrogating the underlying assumptions embedded within the data capturing and processing. The interrogation of the origins of the data would entail consideration of the ethical discourses surrounding the use of learning analytics within the institution.

The Ethical Discourses Surrounding Analytics in Higher Education

The Threat of Algoracy

The rise in the use of big data within institutions occurs amid concerns about privacy, surveillance, the nature of evidence in education and the ethical considerations involved in educational provision (Danaher, 2014; Prinsloo et al., 2015). The overregulation of institution-student interaction through the use of algorithms increases the risk of coming to view students sources of data and passive recipients of services and tuition (Prinsloo et al., 2015). Danaher (2014) posits an extreme case of data application where corporations and governments not only monitor all individual behaviour and interaction but also nudge our decisions in specific directions that are considered as desirable. Morozov (2013) further argues that

we can now be pinged whenever we are about to do something stupid, unhealthy or unsound. We wouldn't necessarily need to know why the action would be wrong; the system's algorithms do the moral calculus on their own. Citizens take on the role of information machines that feed the techno-bureaucratic complex with our data. (para. 13)

The threat lies not only in the invisible external control of our day to day behaviours to meet an undisclosed political or ideological agenda. The additional risk is that decision making structures may become increasingly beholden to the algorithms without necessarily understanding how they work (Danaher, 2014).

Within the higher education context the use of big data to “guide” students into the “appropriate” field of study appears benign on the surface. The process can take a darker turn if the application of the data is used to take away the individual right to decide for themselves on the basis of the outcome provided by the algorithms applied. Students who are informed that their profile is ill suited for a particular degree pathway may be directed into one shown to be more likely to yield successful completion. This decision can be taken without due consideration of the individual freedom to define a chosen career path.

Morozov (2013) outlines his views on three provocative positions in countering the shift toward algorithmic regulation. First, the issue of personal privacy needs to be politicised to promote scrutiny of data intensive problem solving and question the provenance of the data. Furthermore, the ideology

underpinning the algorithm and the political consequences of the decisions made using the algorithm should be brought under scrutiny. Second, a shift in thinking about our preconceptions and ideas about the value of our personal data is necessary to provoke considered negotiation of how much we reveal and to whom.

As such, Morozov (2013) proposes sabotaging the system by resisting the trend toward self-monitoring and short-term monetary gains in exchange for personal data. Third, he proposes the development of provocative digital services designed to nudge citizens to question the hidden political dimensions to each act of data sharing. The irony here lies in the utilization of another algorithm to decide which data is personally valuable and sensitive, creating the potential pitfall that the developers of this service may decide on behalf of students what data they should or should not share.

Learning analytics displays the patterns in the data, the reasons for the patterns remain subject to the underpinning ideology applied to the interpretation of the data. According to Siemens (2005) decisions within higher education are sensitive to the initial conditions when the decision was made. Should the underlying conditions change, then the decisions made are longer correct as they were when the decision was made (Siemens, 2005). This paper shares the position that the defining challenge of our time is inequality.

Education is seen as the best way available to achieve the goal of a populace that is able to improve their position in life and live life in terms of success that is self-defined (Siemens, 2005). However, if the education system is structurally geared toward privileging certain classes of knowledge and people then it may only reinforce existing disparities. In situations where formal education is made a prerequisite to accessing the labour market required for social advance, then education is a means of deepening race and class distinctions (Siemens, 2005).

The social and material inequalities alluded to above can be seen to be structurally embedded within an analytic system where the focus is exclusively on the technical aspects of the system. “An exclusively technical civilization ... is threatened ... by the splitting of human beings into two classes—the social engineers and the inmates of closed social institutions” (Habermas, 1974, p. 274). The statement by Habermas takes on particular relevance in the Unisa context with the emphasis on agency on the part of the institution and the student (Prinsloo et al., 2015).

The implication of acknowledging agency entails integrating students as active contributors to institutional knowledge about the factors affecting performance, engagement, throughput, retention, motivation, locus of control, among others. However, from a data science perspective, the active inclusion of students within the process may have implications for the observer effect. This effect is also known as the Hawthorne effect describes how participants in a research study modify their behaviour to appear to conform to the expectations of the observer.

The risk here is that students may alter their behaviour to conform to what the learning analytics system measures, which may be a proxy for an effect that we are unable to measure directly. For instance, should students be presented with a ranking system for their login activity into the LMS as a proxy from engagement risk, we may observe an increase in LMS activity with little or no impact on performance. Furthermore, there is the additional possibility of students faking good by granting the appearance of engagement by contributing minimally within the LMS. While there are a number of pedagogical and measurement measures that can be taken in mitigation of the altered behavioural

pattern. The point here is simply that we need to critically and continually question our assumptions about the variables used and the link to the underlying construct.

Considerations in Ensuring an Ethical Student Data Harvesting

In addition to Morozov's (2013) three provocative propositions mentioned earlier, Prinsloo et al (2015) propose five theses for ensuring reliable, valid, integrated data across the institution which is used in a framework which has critically considered the meaning of the results of the analysis.

The first thesis focuses on establishing processes for critically question the institutional assumptions about student engagement, retention and success (Prinsloo et al., 2015). The first thesis is particularly important when considering the use of big data to determine access to support within an institution (Prinsloo et al., 2015). The second thesis focuses on the proposition of an integrated data system which is built with both everyday user needs and the needs of real time analytics in mind. According to Prinsloo et al. (2015), most of the systems at Unisa at the enterprise system level and are focused on data capture and managing data input with little integration between the various systems used at the institution.

The third thesis proposes the development of the necessary skills and capabilities within the institution. The development of these skills must take into consideration the shifts required in skillsets to manage data from multiple sources, navigating data sets while eliminating noise and balancing the interpretations of observed patterns with an understanding of the underlying reasons for the observations made (Prinsloo et al., 2015).

The fourth and fifth theses propose a systems evolution that focuses on systematically mapping the data elements required for reporting and analytics in a manner that is relevant to the institutional context and knowledge base (Prinsloo et al., 2015). These two theses are intended to result in a framework that will govern the analytics system from the data through to decision making. Discussions on implementing the propositions made by Morozov (2013) and the theses proposed by Prinsloo et al. (2015) can be discussed under three main topics, namely, politicising the debate on data access, shifting the thinking on the value of personal data and interrogating the origins of the data.

Politicising the Debate Around Data Access

Acknowledging the debate surrounding the access and dissemination of student data as fundamentally political is positioned as one of the first areas in which to transform the application and implementation of data mining within an institution (Prinsloo & Slade, 2014a). In response, I suggest the formation of a multidisciplinary team to develop a data analytics framework which explicitly sets out the institutional assumptions, practices, and ideology underpinning the data mining endeavours undertaken in the institutions name. The team should have the mandate to act as advocates for ethical practices, play the role of educating academics, administrators, and researchers within the institution on the ethical use of student data as well as provide a critical space where interested stakeholders can question the origins, applications, and actionable decisions taken on the basis of data mining within the institution. The team could ideally consist of curriculum specialists, data scientists, ICT developers, policy developers, academic departments, and representatives from the student body at the institution.

A key point on the agenda of the learning analytics task teams within an institution must be the development of data literacy among all users and decision makers using the outputs of the mining

exercise. A second key point would be to interrogate the data collection, processing, and integration practices within the institution to eliminate or mediate the introduction of unknown biases that will negatively affect the quality of the data around key indexes.

A further step in politicising the debate around learning analytics is to take the point of departure that learning analytics is a decision making tool and should not be the decision maker within the system. The final decision on the risk exposure and subsequent interventions should be a dialectical process between the data, faculty and the student since, as described above, there are multiple issues facing the data integrity of the available databases within the Unisa system, issues that are not unique to the institution (Prinsloo et al., 2015).

To allow for the dialectical process, it is therefore important to establish a code of practice that is reflective of the underlying ideological assumptions that underpin teaching and learning at the institution as well as the values of the institution, student and higher education system. Jisc, a UK based research hub for higher education, has established a code of practice based on the following principles:

- Responsibility
- Transparency and consent
- Privacy
- Validity
- Access
- Enabling positive interventions
- Minimising adverse impacts
- Stewardship of data (Sclater, 2015).

The purpose of the code of practice is to ensure that the activities around learning analytics are carried out responsibly by addressing the key ethical, legal, and logistical issues that arise during the implementation of learning analytics within an institution (Sclater, 2015). The principle of responsibility places emphasis on identifying and providing specific responsibility for the use of learning analytics in the institution. The specific responsibilities identified include the collection of the data to be used, the anonymization of the data where appropriate, the analytics to be performed on the data, the interventions carried out and the retention, and stewardship of the data (Sclater, 2015).

These principles, while simple in theory, prove to be quite complex in practice, as stated above, the data sources used for analytics at Unisa emanate from five distinct systems, each with its own processes, procedures, and function, the roles of these systems are operationally focused and the data collected requires transformation prior to use of analytics. The responsibility for the transformation of the data and the legality surrounding the transformation, is an area that requires specific definition of responsibility at the institution, an area not addressed in the Jisc code of practice (Prinsloo et al., 2015).

Transparency and consent refer to making explicit arrangements to alter the existing agreements between students and the institution to access the data for the purpose of analytics. The current approach is to approach students for consent on a project by project basis which, while upholding the principle of transparency and consent, proves to be cumbersome in the development of real time analytics (Prinsloo et al., 2015). Validity requires that institutions facilitate open spaces for robust debate on the validity of the data and analytic processes. This is contrast to the analytics being produced as a black box for consumption of managers, academics, tutors and students by specialist researchers (Sclater, 2015).

Access requires that institutions make the analytics available to students in a meaningful and accessible format (Sclater, 2015). In practice this would entail not only making the results of individual, cohort, and institutional analytics available online but also provide reasonable access to student counselling services to facilitate the interpretation of the results in a way that is conducive to ownership of learning by students (Sclater, 2015).

The debates around enabling positive interventions and minimising adverse impacts should focus on recognising that analytics can never give a complete picture of an individual's learning and may ignore personal circumstances (Sclater, 2015). Furthermore, the debates should hold that norms, trends, categorisation, or labelling do not bias staff, student, or institutional perceptions toward discriminatory attitudes or increase power differentials between student and the institution (Sclater, 2015).

The stewardship of the data sets the requirement that all data storage must comply with the existing legislative frameworks regarding protection of personal information and individual privacy governing the country in which the institution operates (Sclater, 2015). In South Africa, the *Protection of Personal Information Act 4 of 2013* (POPI) in conjunction with the King III Code on Good Governance are cited as the governing legislative principles for data protection (The Republic of South Africa, 2013). POPI takes precedence over any other existing legislation, except where existing legislation provides for more extensive conditions for the lawful processing of personal information than is provided for in POPI (The Republic of South Africa, 2013). Researchers applying data mining techniques have to balance two legislative imperatives. The first is the entrenched right to access to information if the information required for the exercise or protection of a right. The second is the right to protection of personal information as far as the limitation of access is reasonable and justifiable (The Republic of South Africa, 2013).

Within the field of higher education, access to information on student performance, predictions on risk of dropout, and failure all link to the right to access to quality education. The rights listed need to be balanced against the right to privacy and personal data protection (The Republic of South Africa, 2013). The debates around where the ideal balance between these two rights lies must take place within open fora as part of the process of politicising the debate around data access.

Shifting Thinking About Value of Personal Data

One of the simpler methods of ensuring student ownership is to provide an opt in system for personalisation metrics (Morozov, 2013). The opt in system cannot hold hidden costs to non-participation for students such as losing access to interventions that hold potential benefit for the student (Prinsloo & Slade, 2014a). The personalisation aspects of the analytics system should still be

available to students with the option of removing their data once they have completed their field of study (Prinsloo & Slade, 2014b).

Mediated access with access to interpretation support is necessary as the psychological impact of being flagged as a high risk student has been documented by various authors (Prinsloo & Slade, 2014a). Online career counselling, among other interventions, is currently available within Unisa. By leveraging the existing system, it can be possible to provide mediated access to the indexes where students are guided through the interpretation of their particular risk status. Students can also be directed to the interventions they can access to address specific aspects of their unique risks.

Whether this support should be made mandatory for each student who has access to their performance metrics is an issue for debate. On the one hand, the POPI act clearly states that student data remains the property of the individual while the institution is made the legal custodian of the data (The Republic of South Africa, 2013). On the other hand, unsupported access to predictive analysis on future performance could have potentially demoralising and demotivating effects on student performance, particularly for those students who are flagged as high risk (Engle & Tinto, 2008).

Opening student access into data collection process is an attempt at responding to the criticism that students cannot be viewed as static data points but should be acknowledged as individuals with agency. One aspect of partnering with students involves entering into dialogue with students to assist in building a reliable matrix of data. This is an activity that will require that, at the very least, students are allowed to correct and update any erroneous or omitted information used to construct the risk profiles of students at the institution (Morozov, 2013; Prinsloo et al., 2015; Selater, 2015).

One of the risks with using analytics for making predictions is to mistake noise for actual signals (Prinsloo & Slade, 2014). The criticism levelled against the Purdue Course Signals project is a case in point where the claim that the project could provide a 21% increase in retention came under intense scrutiny by Edutech bloggers such as Caulfield (2013) and Feldstein (2013). These authors questioned the causality model of the CS project with regard to the retention figures claimed. According to Feldstein (2013), the central problem with the analysis conducted in the CS project is that it did not control for longitudinal effects when predicting retention.

According to the CS model, students who took two or more CS courses had a higher likelihood of persisting than students who took one CS course or less (Arnold et al., 2012). The analysis by Feldstein (2013) and Essa (2013) showed a reverse causality between the CS programme and student retention which showed that students who persisted in the system took more CS courses. The reasoning allocated to the observed effect was ascribed to the timing of the CS courses which were semester courses and it therefore stands to reason that those students who dropped out within the first semester would have taken less of the CS courses. The claim that the exposure to CS courses increases retention is considered conceptually flawed with regard to retention but the authors concede that the model does account for some performance in the courses but at levels significantly lower than claimed initially (Essa, 2013; Feldstein, 2013).

Interrogating the Origins of the Data

The technical limitations to data mining at Unisa described how the integration of data from systems built for a purpose separate to the goals of analytics may facilitate a situation where data is assimilated

into the analytic process without the necessary interrogation of the origins of the data and the underlying assumptions that informed the original composition and collection of the data.

The use of socio-economic status (SES), gender, race, age, and school performance are often used as contextual predictors of student performance (Jayaprakash et al., 2014; Prinsloo et al., 2015; Subotzky & Prinsloo, 2011). The purpose of the analytics on these particular attributes is often to develop an understanding of how student history and environment can and does impact on performance within a higher education institution (Arnold et al., 2012; Jayaprakash et al., 2014; Prinsloo et al., 2015; Wolff, Zdrahal, & Pantucek, 2013). The position taken in this paper is that the adoption of these variables without interrogating the underpinning assumption about what these variables mean within the South African context increases the risk of mistaking data noise for true correlation or increasing the risk for incorrect causal assumptions as was the case in the CS project.

A case in point is the use of race as a proxy for access to quality of education, access to material resources, and social capital, particularly within the history of the South African context (Broom, 2004; Jansen, 2003; Karlsson, 2010). While I am not implying that the socio-political history of South Africa does not play a role in preparedness for entering and succeeding in higher education, I am contesting the assumption that race is the most accurate proxy for these concepts available.

For instance, the quality of schooling could be obtained by drawing a longitudinal analysis of the various schooling districts within a central database with overall school performance, performance in English as the language of learning and teaching, mathematics, sciences, and accounting as key variables. In addition, drawing on the Department of Basic Education's quintile school system, which aims to classify the relative economic access of schools across five resource quintiles may be used as a proxy for access to quality of schooling.

The quintile system has a number of issues relating to the classification of schools and the permeability of the quintiles for those schools falling within the borders of each category. Nevertheless, using the quintile system may provide a better proxy for access to educational resources than race. My reasoning here is that by using race and home SES as indicators of quality schooling, we assume uniform access to a uniform quality of education among a particular race and SES class.

The issue raised above can be further elaborated to include the institutional assumptions about the ideal student profile and coupled with a critical interrogation of the actual student profile and the extent to which the institution has adapted to the needs of the students (Prinsloo & Slade, 2014b). Only a relatively low proportion of student success is explained by traditional statistical modelling techniques as they only establish valid and reliable relationships between relatively few variables within a specific context (Subotzky & Prinsloo, 2011).

Student success within the higher education landscape is a complex and multi-layered phenomenon that is the outcome of the relationships between different predictor variables particularly across institution types, heterogeneous student profiles, and disciplinary contexts (Subotzky & Prinsloo, 2011). The model suggested by Subotzky and Prinsloo (2011) speaks to the idea of interrogating the assumptions about the ideal relationship between student and institution within the socio-economic context but challenge the idea that the imperative is on the student to assimilate into the institutional culture.

Rather they propose a model that utilises the information mined on students is used to develop an understanding of how the institution can respond to the student needs based on their profile (Subotzky & Prinsloo, 2011). Therefore the range of variables integrated as contextual variables utilised within the Socio-critical model cannot simply focus on the variables situated with the students. The model should also integrate institutionally situated variables such as academic workload, academic experience and qualification, module performance risk, and systemic operations that impact on student performance within the institution (Prinsloo et al., 2015).

Linking back to the principles in the code of practice, the integration of these variables would require that the students have both access and understanding of how the institutionally situated variables impact on their chances of success. This would entail tracking the performance of individual academics, departments, and colleges at the module and programme level through daily activities in a similar manner to which the traditional approach to learning analytics tracks student activities and interaction with the system (Prinsloo, 2009; Prinsloo et al., 2015; Subotzky & Prinsloo, 2011).

Concluding Remarks

In conclusion, creating a cultural shift by debating and negotiating the ideological and value base on which the analytics will be built is one of the first steps in developing a learning analytics system that is both ethical and effective. Creating a shift in thinking about data networks away from institutions as isolated data spheres into a network of publically accessible data that is directly related to the goal of student analytics. As mentioned above, by integrating with the Department of Basic Education databases on school quintiles and longitudinal performance, we could obtain a sense of student performance in context. Creating a code of practice through a dialectical process based on the Socio-critical model of student success at Unisa is an important step in politicising the debate around student data and should focus on issues such as data ownership, student autonomy, and specific accountability for the various aspects of data collection on student information.

Developing the technology that will allow students to track the institutions access to their personal data and manage the level of access that the institution has to personal data at any time as well as the epistemological access to understanding how and why the data is being utilised. Students should also have the option to transfer their data between institutions which link to the point above about rethinking the idea of institutions as isolated information nodes and as part of larger network of information on student performance trends in South Africa.

References

- Arnold, K. E., Hall, Y., Street, S. G., Lafayette, W., & Pistilli, M. D. (2012, May). Course signals at purdue : Using learning analytics to increase student success. *LAK '12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 1, 267–270.
- Broom, Y. (2004). Reading english in multilingual south african primary schools. *International Journal of Bilingual Education and Bilingualism*, 7(6), 506–528.
doi:10.1080/13670050408667828
- Caulfield, M. (2013). What the course signals “kerfuffle” is about, and what it means to you. *EDUCAUSE.edu*. Retrieved from <http://www.educause.edu/blogs/mcaulfield/what-course-signals-kerfuffle-about-and-what-it-means-you>
- Danaher, J. (2014). Philosophical disquisitions: Rule by algorithm? Big data and the threat of algocracy. *Institute for Ethics and Emerging Technologies*. Retrieved from <http://philosophicaldisquisitions.blogspot.com/2014/01/rule-by-algorithm-big-data-and-threat.html>
- Engle, J., & Tinto, V. (2008). Moving beyond access: College success for low-income, first-generation students. *The Pell Institute for the Study of Opportunity in Higher Education*, 1–38.
- Essa, A. (2013). Can we improve retention rates by giving students chocolates? *Analytics, Innovation, Research*. Retrieved from <http://alfredessa.com/2013/10/can-we-improve-retention-rates-by-giving-students-chocolates/>
- Feldstein, M. (2013). Course signals effectiveness data appears to be meaningless (and why you should care). *e-Literate*. Retrieved from <http://mfeldstein.com/course-signals-effectiveness-data-appears-meaningless-care/>
- Habermas, J. (1974). *Theory and practice*. London: Heinemann.
- Jansen, J. (2003). Educational change in South Africa 1994-2003 : Case studies in large-scale education reform. *Education reform and management publication series* (Vol. II). Johannesburg: Joint Education Trust.
- Jayaprakash, S. M., Moody, E. W., Eitel, J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6–47.
- Karlsson, J. (2010). Gender mainstreaming in a South African provincial education department: A transformative shift or technical fix for oppressive gender relations? *Compare: A Journal of Comparative and International Education*, 40(4), 497–514.
doi:10.1080/03057925.2010.490374
- Morozov, E. (2013). Why our privacy problem is a democracy problem in disguise. *MIT Technological review*. Retrieved from <http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/>

- Prinsloo, P. (2009). *Modelling throughput at Unisa: The key to the successful implementation of ODL*. Pretoria: University of South Africa.
- Prinsloo, P., Archer, E., Barnes, G., Chetty, Y., van Zyl, D., & Zyl, V. (2015). Big(ger) data as better data in open distance learning. *International Review of Research in Open and Distributed Learning*, 16(1), 284–306.
- Prinsloo, P., & Slade, S. (2014a). Student data privacy and institutional accountability in an age of surveillance. In *Using Data to Improve Higher Education* (pp. 197–214). Rotterdam: SensePublishers.
- Prinsloo, P., & Slade, S. (2014b). Educational triage in open distance learning: Walking a moral tightrope. *The International Review of Research in Open and Distributed Learning*, 15(4), 306–331.
- Slater, N. (2015). *Code of practice for learning analytics*. London: Jisc. Retrieved from <https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics>
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1), 1–8.
- Simpson, O. (2006). Predicting student success in open and distance learning. *Open Learning*, 21(2), 125–138. doi:10.1080/02680510600713110
- Subotzky, G., & 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. doi:10.1080/01587919.2011.584846
- The Republic of South Africa. Protection of Personal Information (POPI) Act no. 4 of 2013 (2013). Pretoria: The Presidency. Retrieved from <http://www.justice.gov.za/legislation/acts/2013-004.pdf>
- Venter, A., & Prinsloo, P. (2011). The paradox between technology adoption and student success : A case study. *Progressio*, 33(1), 43–62.
- Wolff, A., Zdrahal, Z., & Pantucek, M. (2013). Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *LAK '13 Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 145–149.

