

Use of Analytics in South African Higher Education

The Siyaphumelela project, funded by the Kresge Foundation, supports five South African universities to investigate the use of student data analytics to improve student success. In order to support this initiative *Saide* initiated a project to explore the ethical use of data within a South African context. This included a workshop with these five institutions in 2015, and the development of draft narrative framework for the ethical use of data workshopped and presented at the Siyaphumelela 2017 conference (Appendix A). Professor Paul Prinsloo (Unisa) facilitated both workshops, gave a keynote presentation on, and authored, the framework.

This framework includes a number of guiding principles that support the ethical use of data to support student success:

- The moral relational duty of learning analytics:
Learning analytics should be viewed as relational and emancipatory, rather than a process that perpetuates injustice and inequality.
- Defining student success in the nexus of student, institution and macro-societal agencies and context:
“How can we collaboratively craft a shared understanding of what engagement means for them [students], for us, and then determine what criteria are most useful and descriptive of where they are coming from, where they are and where they want to go?” (p. 24).
- Understanding data as framed and framing”
Data never provide the complete or full picture and that the data we do have of our students are glimpses and often proxies of complex, layered lives” (p. 25).
- Student data sovereignty:
“Student data are not something separate from students’ identities, their histories, their beings ... data are an integral, albeit informational part of students’ being.” (p. 25).
- Accountability:
“[H]igher education institutions are also accountable to students ... [and] learning analytics allows institutions to grow in their understanding of the complexities of learning” (p. 28).
- Transparency:
“[T]ransparency, as a principle in the ethical collection, analysis and use of student data underpins the other principles ... aiming to institutionalise ethical data collection, analysis and use” (p. 32).
- Co-responsibility:
“[E]mphasises the interdependency between institutions and students in facilitating effective and appropriate learning experiences” (p. 33).

For purposes of clarity, the following definitions of analytics and different types of analytics, will form the basis of usage in this document.

Analytics: Statistical techniques are applied to data, usually to answer a particular question, and the results are used for modelling, interpretation, decision support, etc. Data sets are

usually large (e.g. all students taking a particular module; success rate of all students in the institution).

Academic Analytics: ‘Academic analytics (AA) is the improvement of organizational processes, workflows, resource allocation, and institutional measurement through the use of learner, academic, and institutional data’¹. Academic Analytics uses aggregate data from different university systems such as the student information system and these data are collected by an institutional research office for decision-support within the university (for instance, for the Executive) and external reporting (e.g. to DHET through HEMIS).

Learner Analytics: The analysis and interpretation of student data in order to support student success effectively. The data on students – biographical, demographic, academic, learning and institutional – on the university system or collected through a survey instrument are used in order to intervene as early as possible if the profile shows any risk. The analytics can be personalised, help us to learn from the data and help us to predict certain outcomes. Learner analytics can help us to decide on the best course of action for a particular student or group of students. Raw data from different sources combine to form derived variables for the purpose of analysis and reporting.

Learning Analytics: ‘Learning analytics (LA) is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs’². LA is usually dynamic, formative and personalized. One purpose is to improve the learning process. Data on student engagement with their learning and their emerging success are gathered through formative marks data, clickstream data, library usage or any other engagement data in order to intervene as early as possible if trends emerge that might lead to a student not fulfilling his or her academic potential. The results can be used to improve teaching and/ or to evaluate and improve student retention and success interventions. Raw data from different sources combine to form derived variables for the purpose of analysis and reporting.

The introduction of learning analytics to support student success is complex and multifaceted. The New America’s “Predictive Analysis in Higher Education”³ and Jisc’s “Code

¹ Long, P. and Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *Educause Review*. <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education>.

² Siemens, G. (2011). 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, February 27–March 1, 2011, <http://tekri.athabascau.ca/analytics/>.
Lang, C., Siemens, G., Wise, A. and Gašević, D. (2017). (eds). *The Handbook of Learning Analytics*. ISBN: 978-0-9952408-0-3. DOI: 10.18608/hla17 <https://solaresearch.org/hla-17/>.
Journal of Learning Analytics: Supporting and Promoting Learning Analytics Research. A journal started in 2014 by the Society for Learning Analytics Research (SoLAR). <http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/3908>.

³ Ekowo, M. and I. Palmer (2017) Predictive analysis in Higher Education. New America (<https://na-production.s3.amazonaws.com/documents/Predictive-Analytics-GuidingPractices.pdf>)

of practice for learning analytics”⁴ discuss the introduction and ethical use of data for learning analytics from two different contexts. The current document provides an integrated overview of these two documents aligned to the guiding principles. In addition, it attempts to identify the role players, processes and responsibilities to ensure that the introduction of student learning analytics is carried out in a responsible manner and includes an understanding of the ethical and logistical matters. It is understood that South African institutions are legally bound to adhere to the terms of the Protection of Personal Information Act (POPI) that is not discussed here. However, as a baseline, POPI requires that the minimum amount of information possible to the performance of our legal reporting duties and our duty to support student success be requested from students and that it be used only for the purpose for which it was collected. Ethics clearance for any research project using student data should require approval from the Registrar of the university over and about clearance from an Ethics Committee.

Planning

Before embarking on the journey to integrate the use of learning analytics to support student success, it is advisable to set goals, that may be part of the institutional strategic plan. The development of goals/vision should clearly articulate the purpose and outcomes associated with the use of learning analytics. In addition, the development of such a vision statement, or set of goals, should include key stakeholders (for example, institutional management, faculty leadership, institutional researchers, academic members and students). A detailed plan could include processes and management roles related to the following:

- Student consent,
- Communications and transparency,
- Data collection,
- Interventions,
- Mitigation of advert impacts, and
- Development.

Communication/transparency

One way to be transparent in the collection and use of data for learning analytics is to communicate all aspects of associated goals including the potential benefits associated with learning analytics, and the associated processes, policies and potential impacts (both positive and negative).

Student consent

The “Code of practice for learning analytics”⁵ discussed a number of criteria that should be considered in order for students to make informed decisions so that an institution could make use of their data to support their success. These include:

- Institutional objectives for the use of learning analytics, the data required to achieve the objectives, data excluded for analyses and reasons why such data are excluded,

⁴ Jisc (2016) Code of practice for learning analytics – June 2015.
(https://www.jisc.ac.uk/sites/default/files/jd0040_code_of_practice_for_learning_analytics_190515_v1.pdf)

metrics used in the analyses, how the data are interpreted, the algorithms⁵ used during data analysis, and who has access to the data.

- Students should be asked for their consent for institutional use of their data to support their success. Consent could be during the registration period or some time thereafter. It is generally included in the contract that the student signs with the university. Students are entitled to opt out (refuse consent). However, as South African institutions are legally bound to report certain student data to the South African government entities, students may not opt out of the reporting of such data. But, data reported to authorities should be clearly stated.
- Any adverse consequences of giving consent should be made clear to students and the consequences if a student opts out should be clearly explained.

Data collection and use

As data are core to the use of learning analytics, it is necessary to ensure that quality data are accurately recorded, made available to those authorised to view the information, is anonymised and securely stored. However, it is understood that, owing to legal reporting, some student data may not be anonymised. Each of these issues is discussed in more depth.

The collection of data should be accurate and comprehensive¹. In order to support students in a timely manner it is necessary to collect accurate data consistently to identify trends and outlying data points when they occur. However, data should be collected from a variety of sources that might be both quantitative and qualitative in nature. This is important as students are not only described by their performance data but also through their well-being. To ensure the validity of data, institutions should understand inaccuracies in the data, account for missing data, and preclude spurious correlations. In addition, the accurate and comprehensive data collection, and the interpretation of data, should be sensitive and appropriate.

Who should have access to learning analytics data needs to be defined. Students need to understand who has access to their data. Policy should make it clear that students are the owners of the data and have consented to selective access to:

- Those who advise them on their academic progress and well-being;
- Faculty members who support them in improving their performance;
- Institutional and Faculty administrative staff (anonymised data) for reporting and high-level analyses; and
- Institutional members responsible for reporting student data to authorities.

Those who have access to anonymised data should not be able to identify individuals from the metadata or re-identify individuals through the aggregating of multiple data sources, or reverse engineer the anonymisation process. Furthermore, institutions need to protect student data especially when data are shared to, or collected by, outside contracted third parties.

Appropriate controls and back-up processes need to be in place for the “retention and institutional stewardship of all data for and generated by learning analytics.”²

⁵ *Algorithm*: A model containing a set of criteria/ parameters (both numerical and non-numerical) with defined steps/ rules that, when applied to data sets, calculates a result, such as a prediction of success.

Interventions

Acknowledging the lasting, inter-generational effects of colonialism and Apartheid institutions collect, analyse and use student data with the aim of addressing these effects and tensions between ensuring quality, sustainability and success (Prinsloo, 2017, p. 9). One of the mechanisms to address past inequalities is to identify and develop interventions to overcome such inter-generational effects.

Jisc² (p.3) advised that institutions “should specify under which circumstances they believe they should intervene when learning analytics suggests that a student could benefit from additional support”. The scope and outcomes of such an intervention needs to be made clear to the student who might decline the advice.

All interventions, and their potential impacts on student success, should be documented and made available to all students so that they can take control of their own learning.

Through the use of learning analytics an institution might identify a number of interventions that could increase student success. Therefore, funds need to be allocated to interventions identified by institutional leaders, faculty members and students.

Mitigation of adverse impacts

“Data [can] never provide the complete or full picture and that the data we do have of our students are glimpses and often proxies of complex, layered lives” (Prinsloo, 2017, p. 25). Institutions need to ensure that “trends, norms, categorisation or any labelling of students do not bias staff, student or institutional perceptions and behaviours towards them, reinforce discriminatory attitudes or increase social power differentials.”² (p.3). Therefore, interventions and approaches to data analyses should be reviewed to allow students to make decisions based on the analytics results of their data that in no way could be used to impact on their academic progress or well-being. Review of all interventions and their impact on students is necessary to identify, and correct, unintended consequences of analytics.

Development

The successful implementation of learning analytics to support student success requires that the institution provides professional development to institutional researchers to minimise biases, guarantee privacy and protection of data; academic staff members in the use of data to improve their teaching and learning designs and interventions; students in their use of their data to improve their performance; and development of institutional skills in the evaluation of interventions that make use of both qualitative and quantitative methodologies.

Role players

The introduction of learning analytics to support student success includes many role players. However, specific individuals should be tasked with specific responsibilities to support the ethical use of data:

- Collections of data for learning analytics,

- Anonymisation of the data,
- Analytic processes,
- Intervention, and
- Retention and stewardship of data.

The following two sections, adapted from the “Jisc Model Intuition Learning Analytics Policy” and “Jisc Model Student Guide to Learning Analytics” can be used to develop your own policy and student guide.

Model Institutional Learning Analytics Policy⁶

Introduction

1. The collection and use of data about students and their learning are providing new opportunities for institutions to support learners and to enhance educational processes. Learning analytics systems present visualisations of student learning activity and provide predictions of attainment. These will be used at *[institution]* to assist current students in achieving their study goals, and to help us improve our overall provision of education.
2. The institution will use learning analytics to help meet the following strategic objectives: *[e.g. increasing retention and progression, improving attainment]*. These are key elements of the *[Learning and Teaching Strategy / other relevant strategies]*.
3. The *[University]* will ensure that learning analytics are deployed for the benefit of students, with complete transparency about the data that are being captured, processed and used. All activities in this area will comply with the institution's Data Protection Policy/Governance *[link]* to ensure compliance with the South African Protection of Personal Information Act (2013)⁷.

Responsibility

4. Overall responsibility for learning analytics at *[University/College]* is held by *[senior leader responsible for learning analytics]*. Responsibility for relevant areas of activity is allocated as follows:
 - The collection of data to be used for learning analytics - *[e.g. IT Director, head of Institutional Research, those responsible for learning management system, those responsible for the student information system]*
 - The anonymisation or de-identification of data where appropriate - *[e.g. IT Director]*
 - The analytics processes to be performed on the data and their purposes – *[e.g. DVC Learning & Teaching]*
 - The interventions to be carried out on the basis of the analytics – *[e.g. DVC Learning & Teaching]*
 - The retention and stewardship of data used for and generated by learning analytics – *[e.g. Registrar]*
5. Analytics reports or dashboards presented to students are intended to help them understand how their learning is progressing and suggestions may be made as to how they can improve their practices. Students are responsible for assessing how they can best apply any such suggestions to their learning.

Transparency and consent

6. Students are informed about how their data will be processed when they agree to the *[e.g. data processing consent notice / computing regulations]* upon registration. Any

⁶ Sclater, N. 2016. <https://analytics.jiscinvolve.org/wp/files/2016/11/Jisc-Model-Institutional-Learning-Analytics-Policy-v0.1.pdf>

⁷ Government of South Africa. (2013). Protection of personal information Act. Government Gazette, 26 November 2013. <http://www.justice.gov.za/legislation/acts/2013-004.pdf>

research project requiring students to share data other than that already on university systems will require them to sign an informed consent form (e.g. additional surveys, focus group interviews) Data will be collected for learning analytics in compliance with *[these documents]*.

7. The data for learning analytics come from a variety of sources, including the student record system and the institutional learning management system *[insert name]*. The Student Guide to Learning Analytics *[link to document in student section of institutional website]* will clearly specify:
 - The data sources being used for learning analytics
 - The specific purposes for which learning analytics is being used
 - The metrics used, and how the analytics reports or dashboards are produced
 - Who has access to the analytics, and why
 - Guidance on how students can interpret any analysis provided to them
 - The interventions that may be taken on the basis of the analytics
8. Students will be asked for their consent for any automated prompts or suggestions to be sent to them, based on the analytics. These may include emails, SMS messages or app notifications.

Confidentiality

9. Personally identifiable data and analytics on an individual student will be provided only to:
 - The student
 - *[University]* staff members who require the data to support students in their professional capacity
 - Third parties which are processing data on behalf of the institution. In such circumstances the *[University]* will put in place contractual arrangements to ensure that the data are held securely and in compliance with the POPI Act.
 - Other individuals or organisations to whom the student gives specific consent (e.g. through a signed contract with a donor organisation funding the student's studies)
10. *[University]* IT staff will have access to systems and data in order to maintain proper functioning of systems rather than to access any individual's data.

Validity

11. The quality, robustness and validity of the data and analytics processes will be monitored by the *[University]*, which will use its best endeavours to ensure that:
 - Inaccuracies and gaps in the data are understood and minimised
 - The optimum range of data sources to achieve accurate predictions is selected
 - Spurious correlations and conclusions are avoided
 - The algorithms and metrics used for predictive analytics and interventions are valid
 - Learning analytics is seen in its wider context and is combined with other data and approaches as appropriate
 - Data are used in a way that students perceive as being fair to them.

Student access to personal data

12. Mechanisms will be developed to enable students to access their personal data, and the learning analytics performed on it, at any time in a meaningful, accessible format. Students have the obligation to correct any inaccurate personal data held about themselves.
13. Students will also be able to view any metrics derived from their data and any labels attached to them.
14. On occasion, it may be considered that access to the analytics reports may have a negative impact on the student's academic progress or wellbeing. In these cases, they may be withheld from the student. However, if the student requests them, all their personal data and analytics reports will be made available to them.

Interventions

15. A range of interventions may take place with students. The types of intervention and what they are intended to achieve are documented in the Student Guide to Learning Analytics [*link*]. These may include:
 - Prompts or suggestions sent automatically to the student via email, SMS message or mobile app notification (subject to the student's consent)
 - Staff contacting an individual on the basis of the analytics report if it is considered that the student may benefit from additional support
16. Interventions, whether automated or human-mediated, will normally be recorded. The records will be subject to periodic reviews as to their appropriateness and effectiveness.

Minimising adverse impacts

17. The [*University*] recognises that learning analytics cannot present a complete picture of a student's learning and that predictions may not always be accurate.
18. Students will retain autonomy in decision making relating to their learning; the analytics report or dashboards are provided to help inform their own decisions about how and what to learn.

Model Student Guide to Learning Analytics⁸

This guide sits alongside the [University's] Learning Analytics Policy [link]. A key principle of the Policy is to be completely transparent about all aspects of our use of learning analytics. We want you to understand exactly what data is being collected, how they are being processed and what we will be doing with the information. This document provides these details, and will be updated as our use of learning analytics develops.

What is learning analytics?

The “digital footprints” left when students use [LMS name] and other institutional systems can be combined with data such as grades and past academic history to form derived variables that are important in student success. We already collect most of the data required, and use it to review aspects of our courses and manage our use of resources more efficiently. However, the use of these data for learning analytics is new and will provide additional information to you, lecturers, tutors and support staff.

The resulting picture can give you a better idea of how your learning is progressing. It can also help us to understand how we can best support you to meet your goals and achieve your full potential at the [University].

Another use for learning analytics is to make predictions by comparing a learner's patterns of activity and achievement with those of previous groups of students. These can help us to identify those who may be struggling academically. We can then contact them to see if we can help.

How can learning analytics help me?

Many learners, particularly when starting [University], don't have an accurate perception of how their learning is progressing. The **Learning Analytics Student Web site/dashboard** helps you to view information on your learning activities.

Your 'attainment' i.e. marks and grades are displayed, though these are provisional. That means that they must be confirmed by a Faculty Board, and, until confirmed, may change.

Emails or text messages may also be sent to you automatically to suggest additional support or resources that could help your studies. Messages may also be sent to congratulate you on good performance or improvement.

How can learning analytics help my tutor?

The other main way that learning analytics can help you is by giving your tutors better information on your progress. Your tutors may contact you by phone, email or text message to check that you feel you are on track or to arrange a meeting to review your progress. Tutors have access to data about your engagement and attainment that can be used to provide a focus for conversations between the two of you.

⁸Sclater, N. 2016. <https://analytics.jiscinvolve.org/wp/files/2016/11/Jisc-Model-Student-Guide-to-Learning-Analytics-v0.1.pdf>

Tutors or members of support staff may also receive automated alerts about students who might be at risk, helping them to decide if they should make contact. It is important to note that the risk alerts are only diagnostic: the model will always have a degree of inaccuracy, and cannot indicate with certainty what outcomes a student will achieve. It simply helps us to prioritise those students who are most likely to benefit by consultation or additional help.

What data are being collected for learning analytics?

The following data are being used for learning analytics, in compliance with the POPI Act 2013, the University's Data Protection Policy [[link](#)] and Data Consent information [[link](#)]:

- Background information: your name, identifiers used by the [*University*], date of birth, ethnicity, gender, declared disabilities, contact details, whether your parents were in higher education, your socio-economic background, whether you are an overseas student, contact details, and a link to the photo we hold of you
- Details about your programmes, the modules you are taking, and your tutors
- Details of your assessments, marks and grades obtained
- Details of your activity in [*Moodle / Blackboard*]: logins, resources viewed, assessments submitted and graded, and session timeouts

Is my personal information safe and secure?

Your data are subject to strict security procedures in compliance with the POPI Act of 2013. Only those members of staff who have a professional requirement to support you are permitted to view the data and any analytics reports or dashboards about you individually. In particular, your tutors will be able to view data about your engagement, attainment and any predictions made. If you have any concerns about the [*University*] using your data for the purposes of learning analytics, do discuss these with your tutors, lecturers or Deans.

Your data are also combined with data about other students to help us improve our courses and to support students better overall. These data are not individually identifiable. Data are first "de-identified" i.e. fields such as your student ID are encrypted. It is then sent to a central learning records system managed by the [*University's*], and hosted on servers in physically secure, multi-tenanted private cloud arrays.