Applications of Student Data in Higher Education
Issues and Ethical Considerations

Sharon Slade
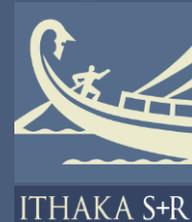
ITHAKA S+R

Ithaka S+R is a strategic consulting and research service provided by ITHAKA, a not-for-profit organization dedicated to helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. Ithaka S+R focuses on the transformation of scholarship and teaching in an online environment, with the goal of identifying the critical issues facing our community and acting as a catalyst for change. JSTOR, a research and learning platform, and Portico, a digital preservation service, are also part of ITHAKA.

Background and Scope

The second Asilomar convention organized by Stanford University and Ithaka S+R in June 2016 brought together a group of academics to facilitate a review of how student data is currently used in higher education. The discussions aimed to synthesize current best practices to specify norms for the ethical use of student data; and inform institutional, national and global policies regarding the research, application, and representation of adult student data.

This paper focuses on the applications strand which sought to yield further insight into:

- The main areas of focus and most promising types of applications for the postsecondary community over the next few years

- A shared understanding of issues (for example, data types or methodologies) that may warrant additional ethical consideration

- The potential for guiding principles which seek to minimize the risks associated with the use of student data to guide or drive their learning.

In particular, the discussion considered both the possibilities and limits of direct intervention in student learning on the basis of data flow, and any risks that should be avoided or at least minimized.

Introduction: The Application of Student Data within Higher Education

Learning analytics is a relatively recent practice, although it builds on the well-established field of educational data mining (amongst others).[1] Perhaps the earliest accepted definition of learning analytics is that it 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]

Early adoption of learning analytics has predominantly occurred in the US, although it is playing an increasing role in determining how many post-secondary education

---

[1] Ryan Baker, "Data Mining for Education," in *International Encyclopedia of Education, 3rd Edition,* eds. Eva Baker, Penelope Peterson, and Barry McGaw (Oxford: Elsevier, 2010), Vol. 7, 112-118.

[2] George Siemens, *Call for Papers for the 1st International Conference on Learning Analytics & Knowledge (LAK 2011)*, 2011, https://tekri.athabascau.ca/analytics/.

institutions (PSEIs) engage with their students at multiple points in the student journey, as well as in the design of teaching and learning content and delivery.

Opportunities and Stakeholders

The EU funded Learning Analytics Community Exchange suggests that the beneficiaries and associated key opportunities offered by learning analytics include:[3]

- *Institutional administrators* in relation to activities such as marketing and recruitment, or efficiency and effectiveness measures.

- *Individual learners* to facilitate a greater understanding of their progress and study behaviors.

- *Teachers and support staff* to inform interventions with individuals and groups;

- *Academic staff* who might wish to adapt existing teaching materials or develop new curricula.

The *NMC Horizon Report: 2015 Higher Education Edition* highlights the increase in the measurement of student learning through data-driven practice and assessment. The report suggests that institutions increasingly seek to gather and analyze "large amounts of detail about individual student interactions in online learning activities. The goal is to build better pedagogies, empower students to take an active part in their learning, target at-risk student populations, and assess factors affecting completion and student success."[4] Others would argue that the datasets involved go beyond those gathered as part of online learning activity, but may include *any or all* data available for collection by the institution, such as that collected as part of the enquiry or registration process, assessment data and data shared by students as part of their daily social and study lives.[5]

The UK-based Higher Education Academy suggests that learning analytics makes available a variety of tools and approaches which provide educators with quantitative

---

[3] LACE, "What are Learning Analytics?" 2015, http://www.laceproject.eu/faqs/learning-analytics/.

[4] Larry Johnson, Samantha Adams Becker, V. Estrada, and A. Freeman, *NMC Horizon Report: 2015 Higher Education Edition*, (Austin, Texas: The New Media Consortium, 2015) 12, http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf.

[5] See Mike Sharples, Anne Adams, Rebecca Ferguson, Mark Gaved, Patrick McAndrew, Bart Rienties, Martin Weller, and Denise Whitelock, *Innovating Pedagogy 2014: Open University Innovation Report 3* (Milton Keynes: The Open University, 2014),
http://www.openuniversity.edu/sites/www.openuniversity.edu/files/The_Open_University_Innovating_Pedagogy_2014_0.pdf.

intelligence to make informed decisions about student learning.[6] Such tools draw upon data from a broad range of sources including behavioral data taken from online learning systems (discussion forums, activity completion, assessments) and functional data taken from student admissions systems and progress reports.[7] As well as basic tracking, a range of interpretative approaches can be applied. These include predictive models (which *infer* information such as drop-out rates and learner outcomes), social network analyses (which examine suggested relationships between networks of individuals, groups and organizations), relationship mining (which analyses links between sets of data patterns such as student success rates), and dashboards (data visualization that enables teachers to give timely feedback to students and integrate results into pedagogical activity).

It may fair to suggest that, although those involved in teaching and learning design and delivery are generally subject experts, they may be less familiar with learning science, pedagogy or educational technology. A draft white paper from the Global Learning Council goes on to say that "intuitive approaches to instructional design often produce poor learning outcomes, whereas research-based approaches have consistently produced significant improvements in learning effectiveness and efficiency."[8] As such, the authors feel that there is a clear responsibility to employ learning analytics to address existing gaps in knowledge and understanding and that doing so *will* lead to enhanced learning outcomes for students.

Possibilities and Purpose

The 2012 UNESCO policy brief on learning analytics confirms the role of learning analytics as one means of optimizing student success.[9] However, it also highlights the role of learning analytics in facilitating key questions around the concept of student success, that is, in defining what it is that higher education is trying to achieve.

In identifying the purposes and range of possibilities offered by collecting, understanding and applying student data then, it seems sensible to first examine the key purpose of post-secondary education institutions themselves. There is perhaps an

---

[6] "What Is Learning Analytics?" Higher Education Academy: Starter Tools, 2015, https://www.heacademy.ac.uk/enhancement/starter-tools/learning-analytics.

[7] Sharples *et al*, *Innovating Pedagogy 2014*.

[8] Global Learning Council, "Technology-Enhanced Learning: Best Practices and Data Sharing in Higher Education," April, 2015, http://globallearningcouncil.org/wp-content/uploads/2015/04/GLC_DRAFT_White_Paper_April_2015.pdf.

[9] Simon Buckingham Shum, "Learning Analytics" (Moscow: UNESCO Institute for Information Technologies in Education, 2012), http://iite.unesco.org/pics/publications/en/files/3214711.pdf.

unspoken presumption that the purposes of higher education are universally agreed and understood. However, most PSEIs will *not* assume a single, simple purpose and, indeed, many operate under a range of policies and guiding principles which serve potentially conflicting needs. Even at the highest level, the purpose of higher education will depend on the (local) context, government policy and cultural and societal expectations. It becomes almost impossible to assign a single purpose to higher education. It may be seen as a pre-requisite for later employment, as a means to facilitate individual growth and satisfaction or focused on the development of high level, intellectual understanding within a particular discipline. For some, education is a means to redress social inequality.

So, in reviewing the application of learning analytics, what are the key issues and areas on which post-secondary education institutions (PSEIs) should focus? A study of learning analytics practice within Australian higher education indicated that senior leaders considered there to be two primary purposes for learning analytics.[10] One cluster largely saw learning analytics as a means to improve student retention, mainly by deploying a technical solution leading to information to prompt action from teachers. The second group saw the potential of learning analytics as a likely disruptor, facilitating the development and improvement of the quality of the student learning experience itself.

So, should the primary goal of PSEIs be to increase qualification or course completion rates of the student body as a whole (to increase the number of students graduating across the piece – that is, the interests of the many outweigh the interests of the few) or should there be a greater focus on improving the learning experience (potentially fewer graduates with a deeper understanding)? Or should there always be an external, societal driver which determines the ways in which a PSEI engages with student data? For example, in the UK, Becca Bland argues that falling completion rates for students from certain disadvantaged backgrounds should be the primary driver for targeting student support.[11]

Participants at the convening were agreed that there is considerable variation of key purpose. These ranged from societal benefit at a large scale ("to transform lives for the benefit of society") and at the individual level ("as a means to social mobility"); to access to educational excellence ("the creation of a contemporary and innovative educational

---

[10] Shane Dawson, Dragan Gasevic, and Tim Rogers, "Student Retention and Learning Analytics: A Snapshot of Australian Practices and a Framework for Advancement," 2016, http://he-analytics.com/wp-content/uploads/SP13_3249_Dawson_Report_2016-3.pdf.

[11] Becca Bland, "To Improve Retention, We Must Focus on Those at Greatest Risk," *Times Higher Education*, April 11, 2016, http://www.timeshighereducation.com/.

experience"); to access for all (through MOOCs, open access, online programs, etc); to preparation for life and work, regional development and supporting the community, etc.

Most were agreed that PSEIs aim to deliver multiple benefits whilst recognizing also that there would be likely differences between public and private providers and differences of focus between various stakeholder groups.

Applications

The range of learning analytics applications in practice will, to a large extent, reflect the presumed purpose of such tools within the (local) higher education context. The recent report from Learning Analytics European Policy (LAEP) project is, *inter alia*, seeking to document evidence of the following types of application:[12]

- General analytics tools – not developed specifically for learning analytics but used in a learning analytics context;

- Learning environment tools – which guide a learning activity, typically informing users who then choose how to act;

- Smart systems – adaptive tools;

- Student-support tools – focusing on student support (retention, completion, etc.) rather than the acquisition of knowledge, skills, or competence;

- Design and planning tools –which support curriculum or learning design, or a related aspect of the environment in which learning is promoted;

A first step perhaps in establishing the range of likely applications is to map against the primary focus of learning analytics.

If the focus is on *retention and completion,* as appears to be the case for many institutions applying analytics, one suite of applications might include the greater use of predictive modelling to determine likelihood of completion prior to beginning study (to tighten the admissions process, for example) or to recommend changes of study direction if things appear not to be going so well; tracking systems could be developed to identify and provide interventions to all students not hitting key milestones, etc. One example of this type of application is Degree Compass developed by Desire2Learn. Degree Compass aims to provide support to students on course selection, reducing enrolment in unnecessary courses and cutting tuition costs. The system identifies those

---

[12] Rebecca Ferguson, "The Implications and Opportunities of Learning Analytics for European Educational Policy (LAEP)," Interim Report, 2016.

students deemed to be at a higher risk for non-completion and provides recommendations as to which courses the students should take in order to complete their degree as well as which courses they are most likely to complete.

On the other hand, if the focus is on the *quality of learning* for the individual student, historical student performance data can be used to better understand how students use course resources as part of their learning; aspects of peer learning; and whether aspects of study are avoided or found to be particularly challenging in order to highlight content for redesign. The NMC *2016 Horizon Report* describes learning analytics as "an educational application of web analytics aimed at learner profiling, a process of gathering and analyzing details of individual student interactions in online learning activities."[13] It places a focus on adaptive learning technologies as a means of analyzing information about students and the ways in which they study to provide a tailored learning opportunity which empowers active learning. A recent white paper by Tyton Partners describes adaptive learning as a "sophisticated, data-driven, and in some cases, nonlinear approach to instruction and remediation, adjusting to a learner's interactions and demonstrated performance level, and subsequently anticipating what types of content and resources learners need at a specific point in time to make progress."[14]

If there is a sense that *students take responsibility for their own learning*, the goal might be to use data to directly provide students with greater insight into (i) their progress against core targets or against other students; (ii) alternative approaches to study which might lead to improved outcomes; (iii) how/where to seek useful study support/resources or (iv) alternatives to the current study choices (biology is not working out so well, have you thought about law as a better match for your academic background/study preferences etc....). Perhaps the best known example of this is Purdue's Signal which reflects back progress to the student and links to specific support, if needed. Another example mentioned in the recent JISC report on current international practice highlights the work undertaken at Open Universities Australia which draws on data from the student profile, learning profile and curriculum profile, and data captured in discussion forums and from open questions to help students plan pathways through

---

[13] Larry Johnson, Samantha Adams Becker, V. Estrada, A. Freeman, and C. Hall, *NMC Horizon Report: 2016 Higher Education Edition*, (Austin, Texas: The New Media Consortium, 2016), http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf.

[14] Tyton Partners," Learning to Adapt: a Case for Accelerating Adaptive Learning in Higher Education," 2016, http://tytonpartners.com/library/learning-to-adapt-2-0-the-evolution-of-adaptive-learning-in-higher-education/.

their modules.[15] Other approaches support the development of social networks and the creation of learning communities.

It is also the case that applications are influenced by *internal or external drivers* which link more closely to institutional success or other external goals.[16] Such applications may include selective admissions (to include or exclude applicants meeting particular characteristics); evaluation of effective advisory support; performance measurement of faculty; effective collaboration with third parties and increased understanding of the validity of proxy measures.

Challenges

There are a number of challenges associated with actively applying analytics to impact on and improve the student outcome. These might be practical constraints, such as affordability, the internal capacity for development and maintenance, availability of appropriate data, etc. or may equally touch upon legal constraints (governed by local or (inter)national legislation) such as privacy law; institutional policy such as a commitment to recruit students from specific demographic groups or broader ethical concerns.

One of the key challenges in understanding how greater engagement with student data might impact both positively and potentially negatively, is in understanding that higher education in itself means a lot of different things. There is no one size fits all approach. In the US, nearly half of all college students attend community colleges; and among those at four-year schools, nearly a quarter attend part time.[17] The focus for education providers will differ enormously.

A report by Randall J. Stiles written for Educause focuses on such challenges and discusses challenges around data governance, including legal data protection requirements, data collection and storage methods, and access to student data.[18] The report also considers data quality and the issues associated with missing, incorrect or

---

[15] JISC, "Learning Analytics in Higher Education: A Review of UK and International Practice," 2016, https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education.

[16] Sharon Slade and Emily Schneider, "Summary of Application Stream of Asilomar II Convening," June 15-17, 2016, http://gsd.su.domains/topics/application/.

[17] National Center for Education Statistics, 2013, http://nces.ed.gov/programs/digest/d13/tables/dt13_303.70.asp.

[18] Randall J. Stiles, "Understanding and Managing the Risks of Analytics in Higher Education: A Guide," EDUCAUSE (June 2012), http://net.educause.edu/ir/library/pdf/EPUB1201.pdf.

misleading data with legal and institutional compliance, the use of third-party systems and issues around ethics and privacy.

Discussions at the Asilomar II convening yielded a number of challenges which include reliance on available rather than appropriate data; (staff) resource limitations; faculty resistance to adoption; data literacy issues and data accessibility.[19]

A recent JISC report flags a number of relevant issues such as the accuracy of predictive modeling and the dangers of unintended consequences, citing the potential loss of students alerted as being potentially at risk who go on to withdraw and who might otherwise have successfully completed their studies without such an intervention.[20] Other related issues around predictive modeling concern the ways in which such models are "trained" on existing data sets and the dangers of inherent algorithmic bias; the limitations of some datasets to fully reflect key issues (for example, VLE engagement data alone can be misleading but is more effective when coupled with other "engagement" datasets and particularly with an evolving record of assessment performance); and the lack of staff expertise in interpreting results and acting upon them in helpful ways. The report also highlights the importance of the contextual integrity of data – information gathered in one context may have less or no relevance if applied within another or when combined with unrelated data from a separate time period, for example. In cases where data is scarce and proxies are applied, false links can potentially highlight issues which do not reflect the student's reality. According to a recent KPMG survey, many colleges lack the internal analytics skills to make best use of existing data.[21]

However sophisticated and accurate the tools are, the ways in which progress information and predicted outcomes are communicated to students is key. Communication can take the form of automated messaging, one-to-one discussion and visual representation via the use of dashboards, for example. Each of these holds its own challenge. Large scale messaging can become bland or impersonal and may be perceived as less relevant (and so ignored). Dashboards can be simplistic or may lack necessary links to next step support. One to one discussions with support staff rely on staff having both the time and key skills to route students toward the best outcome for them.

Whatever the means of communication, any PSEI proactively informing students of progress or potential outcomes needs some awareness of the impact of making an intervention. Even a fairly passive approach, such as an automated email, can trigger a

---

[19] Sharon Slade and Emily Schneider, "Summary of Application Stream of Asilomar II Convening."

[20] JISC, "Learning Analytics in Higher Education: A Review of UK and International Practice."

[21] KPMG, "Embracing Innovation: 2015-2016 Higher Education Industry Outlook Survey," 2016, http://www.kpmg-institutes.com/content/dam/kpmg/governmentinstitute/pdf/2015/he-outlook-2016.pdf.

large increase in workload for support staff. This can require the institution to revisit the issue of purpose and to consider questions around the relative importance of particular student groups – for example, increasing retention for under-represented social groups at the expense of other groups where support resource is constrained.

As well as practical challenges, there are a huge number of ethical challenges. Even if meeting legislative requirements, many students are unaware of or profoundly uncomfortable with their personal data being used to drive educational outcomes.[22] The extent to which students should be actively monitored is also under debate. When does careful tracking become intrusive?

Issues raised at Asilomar II included clearer understanding of inadvertent harm and the primary interests being served; definition of out of scope datasets and access issues; provision of equitable opportunity; the extent to which harm should be accepted as a necessary step toward greater understanding; tensions between finance and student interests; and broader surveillance issues.[23]

A recent paper by Vanessa Scholes proposes several measures that may mitigate some of the ethical issues associated with use of student data in higher education.[24] These involve the transparency of the screening; the static or dynamic nature of the factors used in analytics; the use of statistics specific to individuals, and the distribution of responsibilities between the student and the institution.

We might flag also the need to actively make decisions around more contentious issues such as the allocation of (support) resources for students who are deemed potentially 'high risk'. Is there a point at which it makes less sense to support a student with a high prediction of attrition and what are the ethical implications of doing so? Some institutions are taking this one step further and using analytics to actively cull certain students in order to improve recorded retention figures.[25].

[22] Sharon Slade and Paul Prinsloo, "Student Perspectives on the Use of Their Data: Between Intrusion, Surveillance and Care," *Challenges for Research into Open & Distance Learning: Doing Things Better – Doing Better Things*, 2014, 291–300, http://oro.open.ac.uk/41229/.

[23] Sharon Slade and Emily Schneider, "Summary of Application Stream of Asilomar II Convening."

[24] Vanessa Scholes, "Analytics in Higher Education: The Ethics of Assessing Individuals on Group Risk," (presentation at DEANZ2016 Conference, The University of Waikato, New Zealand, April 17-20, 2016) http://conference.deanz.org.nz/wp-content/uploads/2016/03/DEANZ16-Conference-proceedings11-April.pdf.

[25] Seth Sykes, "Drowning Bunnies or Saving Lives? Putting Data in the Hands of Academic Advisors," *The evolution*, March 28, 2016, http://evolllution.com/technology/tech-tools-and-resources/drowning-bunnies-or-saving-lives-putting-data-in-the-hands-of-academic-advisors/.

Perhaps the most contentious issue when implementing applications of student data is that of consent. Many higher education institutions have implemented an approach to analytics without formally adapting policies relating to the widening use of student data, whilst those that have addressed the issue have tended to focus on a position of pre-registration informed consent.[26]Further clarity around the boundaries of consent (in terms of related actions and data), the meanings of any available opt-out position and the preferences of the student in a world of significant existing tracking would be beneficial.[27]

Developing guiding principles

There have been a number of publications seeking to inform policy and develop guidelines for the use of student data within higher education. In 2014, the US Alliance for Excellent Education published a report outlining the importance of developing a clear understanding of the potential and rationale for learning analytics.[28] From a European perspective, the JISC code of practice focuses on issues of responsibility, transparency and consent, privacy, validity, access, enabling positive interventions, minimizing adverse impacts, and data stewardship.[29] The Open University in the UK implemented a formal policy in 2014 setting out eight principles for the ethical use of student data in learning analytics.[30] More recently, LACE has developed a checklist contains eight action points that should be considered by managers and decision makers when implementing learning analytics projects.[31]

Given the complexity of the challenges flagged above, it can be difficult to create a single set of guidelines to shape the uses of student data. Many of the issues will depend on institutional and/or national context. However, it is worth considering areas for which some guiding principles might be safely established.

---

[26] For example, see Open University, Policy on Ethical use of Student Data for Learning Analytics, 2014, http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf.

[27] Sharon Slade and Emily Schneider, "Summary of Application Stream of Asilomar II Convening."

[28] Alliance for Excellent Education, "Capacity Enablers and Barriers for Learning Analytics," Jun 25, 2014, http://all4ed.org/reports-factsheets/capacity-enablers-and-barriers-for-learning-analytics-implications-for-policy-and-practice/.

[29] JISC, "Code of Practice for Learning Analytics," 2015, https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics.

[30] Open University, Policy on Ethical use of Student Data for Learning Analytics.

[31] LACE, "Ethics and Privacy," 2016, http://www.laceproject.eu/ethics-privacy/.

Relevant questions might be based around issues such as:

- Activities (applied to all or to subsets of students) which should always be off-limits. For example, should physical tracking systems be used to be able to locate students at any time? Is sight of non-study related online activity permissible?

- Similarly, is any data off limits or never relevant?

- Fundamentally, who "owns" student data? Would it be useful to view personal and study data as a commodity to be traded with the student for clearly defined benefits?

- Should all activities/applications come with full disclosure? That is, should students always be told what is being done with their data, what data are being used, what the models predict etc. – or are there any exceptions/downsides to this?

- Related to transparency, should applications only be used if they can be understood by students (for example, predictive modelling based on regression analysis might be considered as complex for many students)? Should data users really understand what they are doing and what the outputs mean (and don't mean)?

- Should there be limits around the confidence levels of predictions used to drive interventions? Should *anything* predictive be sense checked for context by a human first? (and what is the impact of this on scalability of application)?

- Should all students have equal "rights" to interventions or is it acceptable to only apply analytics to key groups (most likely to pass, most likely to fail, particular demographics)

- Are there possible dangers in creating in-house applications (which reinforce a local mind set perhaps) vs. buying in applications (which may not neatly apply to the local context)

- Are there issues around availability of data (access to; completeness of; or the use of one available dataset as a poor proxy for another, etc.)

In moving toward a set of principles, the institutional context and priorities remain relevant and must be considered alongside relevant legislation. Post-secondary education institutions might consider beginning with an environmental scan of existing data sources, applications and policies which may be further supported by reviewing other good practice. Principles can be supported by the establishment of a broad set of cases and examples to aid understanding together with clear guidance for relevant stakeholders.