Return on investment in higher education retention: Systematic focus on actionable information from data analytics

David C. Gibson
Curtin University, Australia

Simon Huband
Curtin University, Australia

Dirk Ifenthaler
Curtin University, Australia

Eric Parkin
Curtin University, Australia

This article describes the human and technical infrastructure analytics capabilities that have evolved at a university in Western Australia, which have been applied to curriculum and learning data with a focus on the return on investment (ROI) of improving retention. The ROI approach has been used to highlight the benefits of further inquiry and action by decision-makers from the classroom level to school and faculty levels. The article will briefly describe the capability developed and methods underpinning continuous on-demand production of analyses and insights aimed to stimulate inquiry and action to improve retention.

Keywords: Return on investment; retention; learning analytics; data analytics

Introduction

Retention is often defined as the process that leads students to remain within the study program and higher education institution in which they enrol and earn a degree (Borgen & Borgen, 2016; Mah, 2016). Retention has been a subject of much discussion and research in Australian higher education since the early 1950's when government policy began to encourage enrolment. The *Higher Education Standards Panel* report of 2017 reviews that history and outlines current concerns including: raising expectations for completion rates, enhancing access to information, transparency and accountability; and improving articulation across the tertiary sector. In addition, the report points out the need for strengthening outreach, providing career advice and support services to assist with completion, creating intermediate qualifications, creating, embedding and sharing innovative practices including international models, and regulating the system for effective and efficient use of government resources (Higher Education Standards Panel, 2017).

The research program described here, situated in a large university in Western Australia, focuses on several of the above-mentioned concerns by calling attention to the human impacts and potential for 'return on investment' (ROI) to stimulate further inquiry and action. By ROI we call attention to the potential of a desired impact in relation to the effort needed to develop a causal intervention such as a new learning experience or an enhancement to an existing one (Psacharopoulos, 2014). In terms of retention at a university, ROI is often summarised as potential tuition retained and as a corollary, attrition as potential revenue lost. But ROI can also be expressed with other costs and benefits, such as *university reputation lost* if students return home unsuccessful and the news spreads by word of mouth to friends and community (Menon, 2014). The plan of the article is to describe information recently shared at a workshop for Heads of School and Unit (Course) Coordinators, which aimed to introduce the current status of and capabilities for data analytics for learning, teaching and curriculum design. That aim well suits the purpose of this article, which is to share information about how the university has recently focused on engaging curriculum leaders in developing their awareness, skills and interests in data-driven decision-making to improved university retention. The article describes the history of the capability build of the human and technical infrastructure, and presents a summary of analysis models as well as approaches to representing findings and its relation to ROI.

Building the human and technical infrastructure

Beginning in 2010, a pilot study showed that behaviours of students in a school of business could be grouped together to better understand the drivers of retention (Deloitte, 2010). The resulting model, termed the *Student Discovery Model* (SDM), utilised a self-organising map methodology (Kohonen, 1990) to create clusters of behaviours that helped analysts discover new relationships, raise additional research questions and test assumptions and hypotheses. For example, the cluster analysis enabled multiple hypothesis testing, since the groups had not been constrained by a single point of view or intervention. This led to a broader understanding of multidimensionality in certain university settings in which retention plays out differently than in others, and which is lost when students are treated as homogenous. The effort was extended in 2013 to the whole university, which involved creating clusters among 52,000 students over a five year period drawing from 15 data systems (e.g., finance, student records, learning management system) and was used to conduct initial exploration of hypotheses as well as to identify correlations that warranted deeper analysis (Gibson & de Freitas, 2015). By 2015, a pilot project in predictive analytics used machine learning to help make the case for the return on investment of building the university's capability in Student Retention Prediction (SRP) (Chai & Gibson, 2015). This effort was partially successful in that machine learning (ML) demonstrated its usefulness, but was unsuccessful in the sense that the target data or measure used by the ML was based on a timeline that was too long for 'student success workers' to make use of the insights during the current semester. In order to develop the capability for near-real time data needed to address this shortcoming, an investment in data architecture simultaneously established how the new exploratory data analytics would interact with managed data systems of the university (see Figure 1).

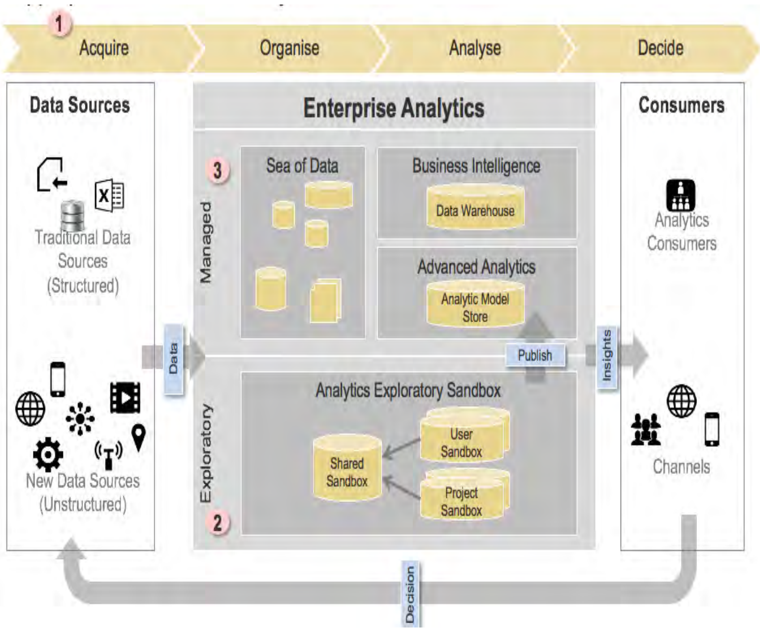

Figure 1: Infrastructure for data analytics includes 1) systems for acquiring, cleaning, organizing and storing; 2) sandbox areas for exploratory analysis; 3) managed data systems for engaging with data consumers

In the planning stage now are tools and processes to engage directly with students based on their own data and to apply the lessons learned to the unit or course level where more dynamic data is produced each semester. To help set the stage for these developments, faculty researchers have been conducting inquiries into the ethics and reactions of staff and students concerning the potential role of data analytics in academic life. An ethical framework has been developed that identifies key questions that require consideration during the process of introducing learning analytics within a university (Roberts, Chang, & Gibson, 2017). Another study explored students' knowledge, attitudes and concerns about big data and learning analytics through focus groups (Roberts, Howell, & Seaman, 2017). Staff registered concerns in a separate study with an overarching concern of coddling and acting in the role of 'helicopter parents' (Howell, Roberts, Seaman, & Gibson, 2018). But despite the challenges, academics saw scope for data analytics to be beneficial if there is collaboration between academics, students, and the university.

Methods, tools and reports underpinning analyses

The methods, tools and reports for accessing data analytics insights for learning, teaching and curriculum design are presented to consumers of university data in terms of products, descriptions, data sources, ease of use and periodic updates (see Table 1). Three primary sources of data are the Learning Management System (LMS), a shared data repository called the 'L Drive' and exploratory data sets created by the Universities Learning and Teaching Unit's learning analytics team. Ease of use reflects whether the user can access and make use of the data product without expert assistance. Updates to data vary depending on the data sources and the complexity of the data product (e.g., nightly, periodically, or on request). Further information and illustrations of the products are offered below.

Table 1: Access, usability and updates profiles of commonly used analytics tools and methods

Product	Description	Data Source	Ease of Use	Updates
Integrated Reports	Available to all unit coordinators within their LMS access. Regular communications to staff highlight use cases such as: item activity, unit access, engagement, contribution and performance, appeals	Blackboard LMS	Easy	Nightly
Disengaged Students List	Enables identification and contacting students who have not been assessing one or more of their LMS units, at key points of the study period (e.g., census, late withdrawal date).	LT Unit analytics	Easy	On request
SDM Retention Data Pack	Per-student Excel retention data, with multiple enhancements (e.g., handling of replacement packages and majors/streams)	L Drive	Difficult	Periodic
Pass Rates	Enables insights into pass rate, withdraw rate and average mark, unit enrolments, for different cohorts.	LT Unit analytics	Easy	On request
Enrolment Trends	Visualizing year-on-year trends	LT Unit analytics	Easy	On request

Integrated reports

Reports integrated with the LMS offer nightly updated views of student engagement with course materials combined with current grade scores and types of work submissions. Interaction totals for each week, with unit features such as accesses, interactions and minutes of access, allow a teacher to see individual student behaviour in one unit compared with average interactions for all other units being studied at the same time. When the engagement data is combined with current grade scores (see Figure 2) then patterns of academic quality emerge indicating that higher engagement correlates with higher grades. An ROI perspective on these aspects might suggest timing and topics for weekly teacher communications to call students' attention to their use of time and energy to improve their learning outcomes.

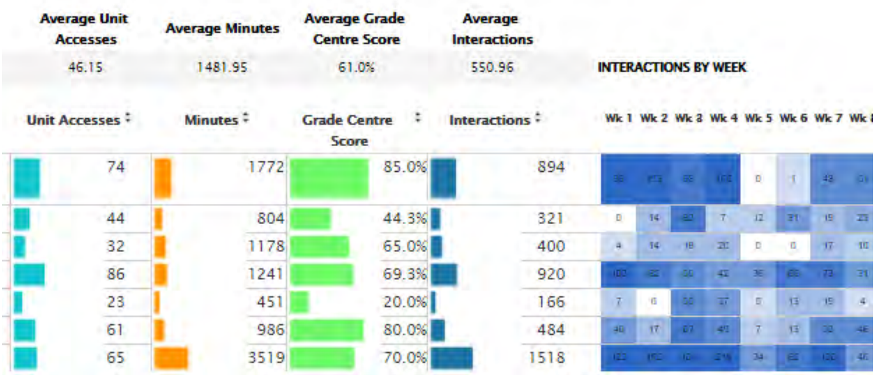

Figure 2: Interactions by week compared with current grade scores

The integrated reports can also be used for curriculum analysis, for example by examining the types of student work being produced at various times of the semester. Student work submissions by week shows peaks of use of academic integrity software in certain weeks and the level of engagement in weekly quizzes and final exams. An ROI perspective on student workload across a whole program might discover that with a shift of a few days, student performance might shift from being a competition among courses to a shifting focus of attention.

Student discovery model retention data pack

The Student Discovery Model (SDM) provides a backdrop for understanding clusters of student behaviours and similarities and also serves as a prepared data source for additional analyses. The preparation steps clean and combine information from 15 sources and place the raw and transformed tables into a production data store used by other data systems, such as for official reporting to the government and tracking the key performance indicators of the university. From the production data store, a data pack is created for each faculty area, with similar visualisation tools and automated analyses that facilitates training and support of decision-makers as well as comparing information and insights (see Figure 3).

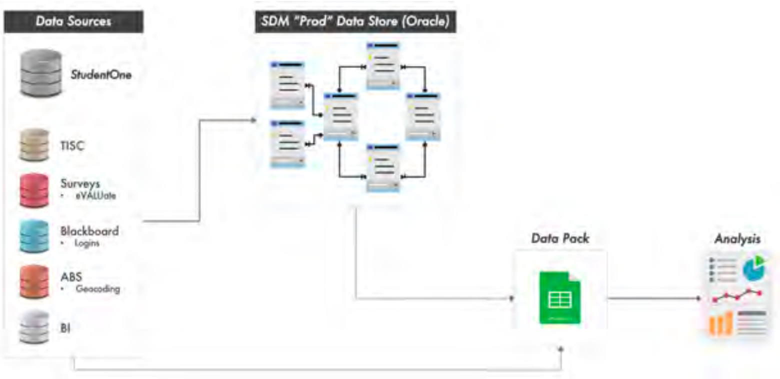

Figure 3: The student discovery model integrates data from 15 sources that flow into a production data store from which a flexible data pack is created for analysis and reporting

An example ROI-oriented product of the SDM created a priority list of programs with an estimated 'Lost Future EFTSL' of that program due to attrition (see Figure 4). In this case, there is a direct impact on school tuition resources that can be estimated as two or more years of lost revenue per student who drops-out in year one, a value estimated in some schools at about $40,000 per student.

University Retention Rate (Headcount)	Possible University Retentions (Headcount)	Lost Future EFTSL
89.5%	19	2.3
73.7%	19	11.1
78.3%	23	14.6
88.9%	18	6.6
100.0%	20	0.0

Figure 4: Retention rates, headcounts and lost future EFTSL (opportunity loss) based on recently historical data

Enrolment trends

Year-on-year comparisons of the dynamic relationship of a unit's enrolment trend with key transition points for attrition provide not only a model of growth (or decline) but also a week-by-week model of time periods when interventions might make a critical difference in retention (see Figure 5). For example, a yearly structural pattern emerges in which rapid drop-outs occur from the date of final enrolment until the census date each year. Sharing and discussing these views of the data helps raise questions about curriculum as well as learning processes. For example, does this pattern occur in all units of the degree program, or only some? Are there non-academic reasons for the pattern? What are the opportunity losses represented by this drop-off pattern? Once the census date has passed, which factors of retention are then most salient? Are there any interventions that might be considered?

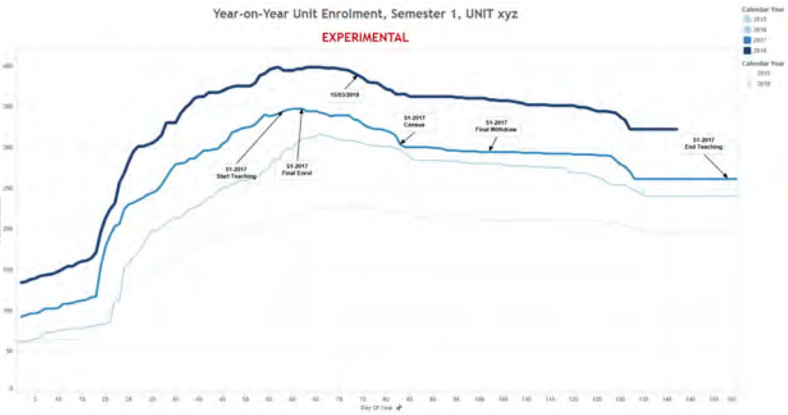

Figure 5: Year-on-year enrolment patterns, showing key transition points

Conclusion

The human and technical infrastructure analytics capabilities of a university can be applied to curriculum and learning data with a focus on a ‘return on investment (ROI) perspective’ for improving retention. The ROI perspective highlights the costs and benefits of data visualisations and analyses for stimulating further inquiry and action by decision-makers at all levels. The production of easy-to-use data sets that can be explored with simple analysis tools has helped build a demand as well as a capability for raising and addressing a wide range of practical research questions across the university. Specific tools and examples are shared here in the hope of initiating professional conversations about data analytics for learning, teaching and curriculum design across universities.

References

Borgen, S. T., & Borgen, N. T. (2016). Student retention in higher education: Folk high schools and educational decisions. *Higher Education, 71*(4), 505–523. doi:10.1007/s10734-015-9921-7

Chai, K., & Gibson, D. (2015). Predicting the risk of attrition for undergraduate students with time based modelling. *Proceedings of the 12th International Conference on Cognition and Exploratory Learning in the Digital Age, CELDA 2015*, 109–116. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84961786619&partnerID=tZOtx3y1

Deloitte. (2010). *Student retention analytics in the Curtin business school*. Bentley, WA.

Gibson, D., & de Freitas, S. (2015). Exploratory Analysis in Learning Analytics. *Technology, Knowledge and Learning*, *21*(1), 5–19. https://doi.org/10.1007/s10758-015-9249-5

Higher Education Standards Panel. (2017). *Improving retention, completion and success in higher education*. Retrieved from https://docs.education.gov.au/system/files/doc/other/final_discussion_paper.pdf

Howell, J. A., Roberts, L. D., Seaman, K., & Gibson, D. C. (2018). Are we on our way to becoming a “helicopter university”? Academics’ views on learning analytics. *Technology, Knowledge and Learning, 23*(1), 1–20. doi:10.1007/s10758-017-9329-9

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*. https://doi.org/10.1109/5.58325

Mah, D.-K. (2016). Learning analytics and digital badges: potential impact on student retention in higher education. *Technology, Knowledge and Learning, 21*(3), 285–305. doi:10.1007/s10758-016-9286-8

Menon, M. E. (2014). Investigating students’ expectations of the economic returns to higher education. In M. E. Menon, Terkla D. G., & P. Gibbs (Eds.), *Using data to improve higher education. Global perspectives on higher education* (pp. 149–163). Rotterdam: Sense Publishers.

Psacharopoulos, G. (2014). The returns to investment in higher education. In M. E. Menon, Terkla D. G., & P. Gibbs (Eds.), *Using data to improve higher education. Global perspectives on higher education* (pp. 121–148). Rotterdam: Sense Publishers. https://doi.org/10.1007/978-94-6209-794-0_8

Roberts, L. D., Chang, V., & Gibson, D. (2017). Ethical Considerations in Adopting a University- and System-Wide Approach to Data and Learning Analytics. In *Big Data and Learning Analytics in Higher Education* (pp. 89–108). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5_7

Roberts, L. D., Howell, J. A., & Seaman, K. (2017). Give me a customizable dashboard: personalized learning analytics dashboards in higher education. *Technology, Knowledge and Learning, 22*(3), 317–333. doi:10.1007/s10758-017-9316-1

Please cite as: Gibson, D., Huband, S., Ifenthaler, D. & Parkin, E. (2018). Return on investment in higher education retention: Systematic focus on actionable information from data analytics. In M. Campbell, J. Willems, C. Adachi, D. Blake, I. Doherty, S. Krishnan, S. Macfarlane, L. Ngo, M. O'Donnell, S. Palmer, L. Riddell, I. Story, H. Suri & J. Tai (Eds.), Open Oceans: Learning without borders. Proceedings ASCILITE 2018 Geelong (pp. 374-379). https://doi.org/10.14742/apubs.2018.1964