Using Predictive Analytics to Support Students and Reduce Attrition: A Rapid Evidence Assessment

Contents

Executive Summary

This paper advocates for a data-driven, proactive approach to identifying students at risk of disengagement and/or discontinuation. It aims to provide the evidence base for University of Westminster's adoption of a predictive learner analytics model. Student disengagement is a complex, multi-faceted phenomenon influenced by multiple factors spanning academic performance, behavioural patterns, attitudinal factors, demographics, and institutional dynamics. Utilising a Rapid Evidence Assessment (REA) methodology, 56 peer-reviewed and industry studies and reports were analysed. This exercise identified 48 (dis)engagement indicators that should be considered when thinking about adopting a predictive model. Additionally, the REA highlighted numerous Machine Learning (ML) techniques used by Higher Education institutions to track student engagement and proactively ensure retention. These were discussed with Random Forest being highlighted as a precise and accurate technique that has been used by our UK HE peers. The paper also considered the importance of considering ethical and operational issues, from data architecture and governance to stakeholder buy-in, ethics, as well as data protection and privacy. Based on this, the following recommendations were made:

1. To foster institutional buy-in and improve data literacy among faculty and colleagues. This will involve detailed stakeholder analysis for all project stages, from conception to completion. Careful consideration should be given to each team's or department's interests, the benefits of adopting a predictive approach being clearly stated.
2. To work towards UoW's technical readiness for data integration and robust ethical governance frameworks are in place. This will be a large piece of work involving colleagues from across the university. Additionally, as the predictive model evolves and new data points are incorporated into it, this will not be a one-off exercise.
3. To improve data literacy among faculty, colleagues, and students. Teaching and professional services staff must be aware of and confident in interpreting the output from the predictive model; the output could take the form of data dashboards. There also needs to be consideration about whether students have permission to view their personal metrics, risk scores, etc. Whatever the final decision, clear rules must be laid concerning the appropriate usage of model outputs.
4. The factoring in of the engagement indicators identified in this paper as well as institution-specific data must be considered when procuring a predictive learner analytics platform.
5. To utilise the predictive insights to identify students at risk of disengagement/discontinuation and colleague knowledge and expertise to design new or deliver existing interventions. Support should be front-loaded as first-year undergraduate students are at particular risk of disengagement and discontinuation.

By leveraging predictive learner analytics, UoW can transition from reactive to proactive strategies, address the signs of disengagement early and enhance student attainment and wellbeing.

Introduction

A 2023 report on the drivers of student (dis)engagement and (dis)continuation at the University of Westminster (UoW) recommended that the set of indicators currently used to identify disengaged students be reviewed.[1] This study constitutes the foundational rationale and blueprint for a larger piece of work rethinking and restructuring UoW's approach to reducing student attrition and improving student wellbeing and academic performance. This paper argues for the utilisation of predictive learner analytics (predictive LA) to *assist* in identification of students at risk of disengaging from their studies. Predictive models can endow Higher Education (HE) providers with a wealth of insight regarding student recruitment and retention by analysing historical and real-time learner data to identify patterns and forecast outcomes. A growing number of HE providers are using data analytics to enhance student outcomes driven by advances in artificial intelligence (AI) and Machine Learning (ML) and exploiting the data-rich environments created by online learning platforms and learning management systems (LMS).[2] Predictive models must be built on apposite foundations to make accurate and meaningful predictions, making calculations based on the most pertinent indicators. Crucially, there must always be human input and deliberation about the most appropriate action based on predictive LA output and each individual student. This is a collaboration between man and machine, the balance between which needs to be carefully and continuously navigated.

This Rapid Evidence Assessment (REA) aims to establish an evidence base which the UoW can draw upon in its deliberations and procurement of a predictive LA model. It does so by reviewing recent research on drivers of student attrition. More specifically, the project aims to 1) determine and tabulate the variables associated with (dis)engagement/(dis)continuation/retention from the wider literature and Westminster-based sources and data, and 2) sketch out popular predictive models and how their outputs have been used to inform interventions and strategies aimed at tackling student attrition. It is clear from the review that attrition is a multi-factorial phenomenon, something which predictive models can be designed to handle. The paper opens with a discussion and definition of key terms used throughout. This is followed by a breakdown of the REA method, detailing the inclusion/exclusion

[1] Scott Rawlinson and Jo Alexander, "Perspectives on Student (Dis)Engagement and Continuation (Academic and Support Colleagues)," Westminster: University of Westminster – Institutional Research, 2023. https://research.westminster.ac.uk/file/1913ce0d2ddd94ee78da7511d2ef03d0f55aedb9b1ff97535503e7c4aacaae6c/372336/Perspectives%20on%20Student%20Engagement%20and%20Continuation%20-%20Academic%20and%20Support%20Colleagues.pdf. This article should also be consulted for an explanation of the current approach used at Westminster to identify disengaged students, 12-14.

[2] Carly Foster and Peter Francis, "A Systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes," *Assessment and Evaluation in Higher Education* 45, no.6 (2020): 822-841; and Carolina Guzmán-Valenzuela, Carolina Gómez-González, Andrés Rojas-Murphy Tagle and Alejandro Lorca-Vyhmeister, "Learning analytics in higher education: a preponderance of analytics but very little learning," *International Journal of Educational Technology in Higher Education* 18, no.23 (2021): 1-19.

protocol, data analysis and write-up. The bulk of the paper consists of the evidence assessment, which resulted in a comprehensive list of indicators of (dis)engagement/(dis)continuation. The assessment identified 48 indicators encompassing student demographic, academic, attitudinal, behavioural characteristics, and institutional factors. This is followed by a discussion of commonly used statistical models concerning the design of predictive LA models, with a particular emphasis on their accuracy and precision regarding the identification of students at risk of disengagement. The penultimate section discusses how disengagement and non-continuation risks have been addressed based on the output of predictive models and how UoW might approach this. The paper concludes with some concluding remarks and recommendations regarding the next steps.

1.0 Definition of Key Terms

1.1 Student Retention, Attrition, and (Dis)Engagement

Clear definitions and assumptions must be stated and justified to identify appropriate indicators and measures of retention, attrition, discontinuation, and disengagement. Retention is our dependent variable; this exercise aims to build an index/directory of independent variables that, through their interconnections, reveal a detailed network of the drivers of attrition, identifying those students presenting the highest risk. AdvanceHE state that retention is about 'students remaining in one HE institution and completing a programme of study'.[3] Retention can be considered a binary variable – students either continue or cease their studies.[4] As this implies, a relationship exists between retention and attrition/discontinuation, with the latter referring to those who leave a programme of study before its completion.

In the period before the cessation of studies, students may become increasingly disengaged with their education. Broadly speaking, student disengagement can be understood as a situation where a student is no longer focused on or feels disconnected from their studies. Academically, student disengagement can be defined as the lack of interest or motivation a student demonstrates towards the learning process. Scratching beneath the surface, (dis)engagement is a fluid state shaped by a network of factors, including student behavioural, attitudinal, demographic, academic, institutional, and extra-institutional context (i.e., political

[3] AdvanceHE, "Student retention and success in higher education," *AdvanceHE*. [Accessed 02/08/2024]. https://www.advance-he.ac.uk/guidance/teaching-and-learning/student-retention-and-success.

[4] Shane Dawson, Jelena Jovanovic, Dragan Gašević, and Abelardo Pardo, "From prediction to impact: Evaluation of a learning analytics retention program," in: *Proceedings of the seventh international learning analytics & knowledge conference*, pp. 474-478, 2017; and Isabelle Archambault, Michel Janosz, Elizabeth Oliver and Véronique Dupéré, "Student Engagement and School Dropout: Theories, Evidence, and Future Directions," in *Handbook of Research on Student Engagement*, eds. A.L. Reschly and S.L. Christenson (Springer, 2002): 331-355.

factors), among others.[5] Furthermore, there is no single direction of travel. Thus, poor academic performance can negatively impact self-perception, leading to discontinuation; conversely, poor self-perception may have a detrimental effect on academic performance.[6] As such, (dis)engagement is shifting ground and requires a holistic approach to be understood. Predictive LA tools can be utilised as an early warning system to identify disengagement before it becomes discontinuation.[7] This discussion has demonstrated familial ties between retention, attrition, (dis)continuation and (dis)engagement. The indicators of student (dis)engagement and (dis)continuation tabulated below (Figure 3) were drawn from this family of literature.

1.2 Educational and Predictive Learner Analytics

Educational LA involves collecting, analysing and reporting data about learners and their environments. The primary aim of educational LA is to optimise learning and the contexts in which it occurs.[8] This data is 'underpinned' by pulling through additional data from university systems and making that information available to relevant colleagues via dashboards.[9] This information can be used by academic and professional services staff to identify students at risk of withdrawing and prompt intervention to prevent this outcome.[10] The field has experienced rapid growth since 2011, with an increasing volume of publications, techniques, methods and applications being presented.[11] The capabilities and appetite for LAs and "big data" solutions have increased in the context of digital transformation, the expansion of online learner environments, particularly during the COVID-19 pandemic, and the shift towards completely

---

[5] Ella R. Kahu, "Framing student engagement in higher education," *Studies in Higher Education* 38, no.5 (2013): 758-773; Megan Louise Pedler, Royce Willis and Johanna Elizabeth Nieuwoudt. "A sense belonging at university: student retention, motivation and enjoyment." Journal of Further and Higher Education 46, no.3 (2022): 397-408.

[6] Dawson *et al*., "From prediction to impact."

[7] Archambault *et al*. "Student Engagement and School Dropout"; Pam Arroway, Glenda Morgan, Molly O'Keefe and Ronald Yanosky, *Learning Analytics in Higher Education*, Research report. Louisville, CO: ECAR, March 2016; Colin Beer, Celeste Lawson, Gemma Mann and Damien Clark, "Measuring engagement: An institution-wide implementation of learning analytics to increase retention," CQUniversity. Conference contribution, 2017; Henk Huijser, Deborah West and David Heath, "The Potential of Learning Analytics to Systematically Address Diverse Learning Needs and Improve Student Retention in Australian Higher Education," *Advances in Scholarship of Teaching and Learning* 13, no.1 (2016): 1-19; Chunping Li, Nicole Herbert, Soonja Yeom and James Montgomery, "Retention Factors in STEM Education Identified Using Learning Analytics: A Systematic Review." *Education Sciences* 12, no.781 (2022): 1-18; Catarina Félix de Oliveira, Sónia Rolland Sobral, Maria João Ferreira and Fernando Moreira, "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education: A Systematic Review," *Big Data and Cognitive Computing* 5, no.64 (2021): 1-33; Dalia Abdulkareem Shafiq, Mohsen Marjani, Riyaz Ahamed Ariyaluran Habeeb, and David Asirvatham, "Student retention using educational data mining and predictive analytics: a systematic literature review," *IEEE Access* 10 (2022): 72480-72503; Nisha S. Raj and V.G. Renumol, "Early prediction of student engagement in virtual learning environments using machine learning techniques," *E-Learning and Digital Media* 19, no.6 (2022): 537-554; and, Anders Larrabee Sønderlund, Emily Hughes and Joanne Smith, "The efficacy of learning analytics interventions in higher education: A systematic review," *British Journal of Educational Technology* 50, no.5 (2019): 2594-2618.

[8] Christothea Herodotou, Galina Naydenova, Avi Boroowa, Alison Gilmour and Bart Rienties, "How Can Predictive Learning Analytics and Motivational Interventions Increase Student Retention and Enhance Administrative Support in Distance Learning?" *Journal of Learning Analytics* 7, no.2 (2020): 72-83; and Lap-Kei Lee, Simon K.S. Cheung and Lam-For Kwok, "Learning analytics: current trends and innovative practices," *Journal of Computers in Education* 7, no.1 (2020): 1-6.

[9] Office for Students, "Suicide prevention and data analytics," *Office for Students*. [Accessed 25/07/2024]. https://www.officeforstudents.org.uk/for-providers/equality-of-opportunity/effective-practice/suicide-prevention-and-data-analytics/.

[10] TASO (Transforming Access and Student Outcomes in Higher Education), *Using learning analytics to prompt student support interventions: Findings from two randomised controlled trials* (TASO: February 2024).

[11] Lee, Cheung and Kwok, "Learning analytics."

online or blended courses. LA tools at UoW and other HE providers fall into the category of *descriptive* LA. While predictive models enable forecasting, early prevention and intervention and are proactive, descriptive tools are reactive and capture a moment in time.[12]

It has been argued that '[t]he purpose of employing analytics in education is to explore trends and patterns using numerous amounts of historical data to predict the future of students' success'.[13] One growing branch of LA is concerned with identifying the data points associated with performance, retention, and wellbeing, among others, and feeding that information into predictive models using ML, often referred to as predictive LA.[14] In terms of retention, predictive LA techniques are applied to identify commonalities among discontinued students from past cohorts, using the patterns to pinpoint current students at risk of disengaging and/or terminating their studies. The utilisation of such techniques marks a move away from reactive to proactive approaches to student retention.

2.0 Methodology and Method: Rapid Evidence Assessment

Methodological choices were driven by the aim of identifying indicators of student (dis)engagement. This paper utilises REA as the existing literature provides an invaluable foundation upon which UoW could develop bespoke predictive LA tools. Prior to the commencement of the review, a protocol (Figure 1) was elaborated detailing the inclusion/exclusion criteria, databases searched and search terms, and how the extracted data was analysed.

Figure 1. REA Protocol

Item	Detail
Inclusion criteria:	• English language. • Full article, report, etc., available. • Published during or after 2010. • Use qualitative, quantitative or mixed research methods. • Peer-reviewed research, institutional research reports conducted by HE providers, and research by organisations working within the HE space (i.e., Higher Education Policy Institute (HEPI), AdvanceHE). • Grey literature, including preprints from open-access repositories such as arXiv. • Paper discusses predictive and/or prescriptive learner analytics, student retention, disengagement from HE, or drivers of discontinuation.
Exclusion criteria:	• Articles, reports, etc., were excluded where only an abstract was available. • Research related to primary or secondary/high school. • Article discussed retention in the context of post-study employment. • Blogs. • Published before 2010.

[12] Teo Susnjak, Gomathy Suganya Ramaswami and Anuradha Mathrani, "Learning analytics dashboard: a tool for providing actionable insights to learner," *International Journal of Educational Technology in Higher Education* 19, no.12 (2022): 1-23.
[13] Shafiq *et al.*, "Student retention using educational data mining and predictive analytics," 72481.
[14] Nabila Sghir, Amina Adadi and Mohammed Lahmer, "Recent advances in Predictive Learning Analytics: A decade systematic review," *Education and Information Technologies* 28 (2023): 8299-8333.

Databases searched and search terms:	• Articles were sourced from PubMed and Google Scholar. • Search terms used included: "student retention AND higher education"; "dropout AND higher education"; "university AND dropout"; "dropout OR retention AND higher education"; "educational analytics AND retention"; "predictive learner analytics AND student engagement OR retention"; "student engagement AND higher education".
Extraction and data analyses:	• Articles were downloaded and saved on a secure OneDrive. • Initial screening of articles involved reading titles and abstracts to establish relevance, followed by a full-text review (see Inclusion criteria). • Key details were extracted from articles, including author(s), country of origin, methodological details (i.e., sample size, methodology and methods, etc.), key arguments and findings. • Data was synthesised into a narrative that discussed indicators of disengagement and tabulated them.

As part of the REA, 56 unique papers were analysed, detailed in Figure 2. The extracted information enabled the following research questions to be answered. Firstly, what are the strongest indicators of student (dis)engagement from HE? Secondly, how can predictive LA models be designed and utilised to address the issue of student retention?

Figure 2. Literature details

ID	Author	Title	Research Methods	Sample Size / References	Country of origin	Year
[1]	Aina *et al*.	The determinants of university dropout: a review of the socio-economic literature	Literature Review	185 papers	Italy, Netherlands, Germany	2021
[2]	Al-Tameemi *et al*.	Predictive learning analytics in higher education: factors, methods and challenges	Literature review	63 papers	UK and Iraq	2020
[3]	Archambault *et al*.	Student engagement and school dropout: theories, evidence, and future directions	Literature Review	110 papers	Canada	2022
[4]	Arroway *et al*.	Learning analytics in higher education	Survey containing quantitative and qualitative items	245 institutions	USA	2016
[5]	Asai	Race matters	Commentary	NA	USA	2020
[6]	Aulck *et al*.	Predicting student dropout in higher education	Quantitative	32,500 students	USA	2017
[7]	Beer *et al*.	Measuring engagement: an institution-wide implementation of learning analytics to increase retention	Review paper discussing development of learning analytics system	NA	Australia	2016
[8]	Calvert	Developing a model and applications for probabilities of student success: a case study of predictive analytics	Case study	NA	UK	2014
[9]	Cochran *et al*.	The role of student characteristics in predicting retention in online courses	Quantitative	2,314 students	USA	2014

[10]	Colvin *et al*.	Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement	Qualitative	32 senior institutional leaders	Australia	2015
[11]	Crane *et al*.	Come out, get out: relations among sexual minority identification, microaggressions, and retention in higher education	Quantitative	152 students	USA	2022
[12]	Dawson *et al*.	From prediction to impact: Evaluation of a learning analytics retention program	Quantitative	11,160 students	Australia, Serbia, UK	2017
[13]	Denaro *et al*.	Identifying systematic inequity in higher education and opportunities for improvement	Quantitative	4,644 individual undergraduate course sections	USA	2022
[14]	Fan *et al*.	Supporting engagement and retention of online and blended-learning students: A qualitative study from an Australian university	Qualitative	Interviews with 41 students in online and blended learning courses	Australia	2022
[15]	Faridhan, Loch and Walker	Improving retention in first-year mathematics using learning analytics	Literature review	18 papers	Australia	2013
[16]	Foster and Francis	A systematic review on the deployment and effectiveness of data analytics	Systematic Review	34 papers	UK	2019
[17]	Freitas *et al*.	Foundations of dynamic learning analytics: using university student data to increase retention	Paper proposes a learning analytics model	NA	Australia	2015
[18]	Guzmán-Valenzuela *et al*.	Learning analytics in higher education: a preponderance of analytics but very little learning	Literature review	385 papers	Chile	2021
[19]	Herodotou *et al*.	How can predictive learning analytics and motivational interventions increase student retention and enhance administrative support in distance education	Quantitative	630 students	UK	2020
[20]	Higher Education Funding Council for Wales	Use of data to support student engagement in higher education	Qualitative	9 case studies	Wales	2024
[21]	Huijser, West and Heath	The potential of learning analytics to systematically address diverse learning needs and improve student	Project report (mixed methods)	NA	Australia	2016

		retention in Australian higher education				
[22]	Jia and Maloney	Using predictive modelling to identify students at risk of poor university outcomes	Quantitative	15,833 students, 88,464 course-specific observations	New Zealand	2015
[23]	Jüttler	Predicting economics student retention in higher education: the effects of students' economic competencies at the end of upper secondary school on their intention to leave their studies in economics	Mixed methods	538 students	Switzerland	2020
[24]	Kahu	Framing student engagement in higher education	Qualitative	NA	New Zealand	2013
[25]	Keane	The Office for Students mental health analytics project, an evaluation	Quantitative	NA	UK	2024
[26]	Lee, Cheung and Kwok	Learning analytics: current trends and innovative practices	Quantitative	24 case studies	China	2020
[27]	Li, Chunping *et al*.	Retention factors in STEM education identified using learning analytics: a systematic review	Systematic Review	59 papers	Australia	2022
[28]	Marôco *et al*.	Predictors of academic efficacy and dropout in university students: Can engagement suppress burnout?	Quantitative	4,061 students	Cross-national	2020
[29]	Matz *et al*.	Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics	Quantitative	50,095 students	USA	2023
[30]	Nurmalitasari, Long and Noor	Factors influencing dropout students in higher education	Mixed methods	108 students	Malaysia, Indonesia	2023
[31]	Oliveira *et al*.	How does learning analytics contribute to prevent students' dropout in higher education: a systematic literature review	Systematic Review	50 papers	Portugal	2021
[32]	Ortiz-Martínez	Analysis of the retention of women in higher education STEM programs	Mixed methods (historical data analysis and survey)	49 survey respondents	Mexico	2023
[33]	Parkes *et al*.	Being more human: rooting learning analytics through resistance and reconnection with the	Qualitative	NA	UK	2020

		values of higher education				
[34]	Paterson and Guerrero	Predictive analytics in education: considerations in predicting versus explaining college student retention	Quantitative	2,352	USA	2023
[35]	Paura and Arhipova	Cause analysis of students' dropout rate in higher education study program	Quantitative	677	Latvia	2014
[36]	Peck	Student analytics: a core specification for engagement and wellbeing analytics	Project report	NA	UK	2023
[37]	Pedler *et al*.	A sense belonging at university: student retention, motivation and enjoyment	Mixed methods (questionnaire)	578	Australia	2022
[38]	Raj and Renumol	Early prediction of student engagement in virtual learning environments using machine learning techniques	Quantitative	7,775	India (using Open University data)	2022
[39]	Rawlinson	Academic and practical information seeking behaviours and needs of international students at pre-arrival and arrived (first year) stages	Qualitative (focus groups)	15 students	UK	2023
[40]	Rawlinson and Alexander	Perspectives on student (dis)engagement and continuation (academic and support colleagues)	Qualitative (semi-structured interviews)	19 academic and support staff	UK	2023
[41]	Rawlinson	'Little Islands': challenges and opportunities for student carers in higher education	Qualitative (semi-structured interviews)	10 students	UK	2024
[42]	Realinho	Predicting student dropout and academic success	Quantitative	4,424 records	Portugal	2022
[43]	Rodríguez-Muñiz *et al*.	Dropout and transfer paths: what are the risky profiles when analysing university persistence with machine learning techniques	Quantitative	1,055 students	Spain	2019
[44]	Rotar	A missing theoretical element of online higher education student attrition, retention, and progress: a systematic literature review	Systematic Review	30 papers	Russia	2022
[45]	Sclater and Mullan	Learning analytics and student success	Literature Review	14	UK	2017
[46]	Seidel and Kutieleh	Using predictive analytics to target and improve first year student attrition	Project report (quantitative)	NA	Australia	2017

[47]	Sghir, Adadi and Lahmer	Recent advances in predictive learning analytics: a decade systematic review (2012-2022)	Systematic Review	74 papers	Morocco	2022
[48]	Shafiq *et al*.	Student retention using educational data mining and predictive analytics: a systematic literature review	Systematic Review	100 papers	Malaysia	2022
[49]	Shaikh and Asif	Persistence and dropout in higher education: review and categorisation of factors	Systematic Review	76 papers	Pakistan	2022
[50]	Sønderlund, Hughes and Smith	The efficacy of learning analytics interventions in higher education: a systematic review	Systematic Review	11 papers	UK	2019
[51]	Stylianou and Milidis	The socioeconomic determinants of university dropouts: The case of Greece	Quantitative	1,120	Greece	2024
[52]	Susnjak, Ramaswami and Mathrani	Learning analytics dashboard: a tool for providing actionable insights to learners	Systematic Review	17	New Zealand	2022
[53]	TASO	Using learning analytics to prompt student support interventions	Project report (quantitative)	Report on two randomised control trials	UK	2024
[54]	West	Learning analytics: assisting universities with student retention	Project report (mixed methods)	NA	Australia	2015
[55]	Wild	Trajectories of subject-interests development and influence factors in higher education	Quantitative	4,345	Germany	2022
[56]	Wolff *et al*.	Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment	Quantitative	7,701	UK, Germany, Czechia	2013

3.0 Rapid Evidence Assessment

The REA highlighted multiple drivers of student attrition, which can be divided thematically: 1) student demographics (i.e., sex/gender, ethnicity, age, etc.), 2) student attitudinal and academic factors (i.e., wellbeing, motivation, assessment and exam scores, etc.), 3) student behavioural factors (i.e., engagement with virtual learning environment (VLE), etc.), and 4) institutional factors (i.e., perceptions of teaching quality; course content (e.g., relevance, quality, etc.)). This review discusses each of these themes sequentially. However, it is recognised that disengagement is a complex, multi-factorial phenomenon, further reinforcing the value of models capable of handling this complexity. A moment should be taken to explain the referencing system adopted in section 3.5. The first column of Figure 2 consists of an ID number (1-56). To avoid the footnotes at the bottom of each page becoming unwieldy, when reference is made to an indicator of disengagement in Figure 4, the ID number of the appropriate paper appears next to the indicator in the following format: [1]. Where multiple papers are referenced, ID numbers appear in square brackets separated by a comma, i.e., [1, 2, 3].

3.1 Student Attitudinal and Academic Factors

Understanding the evolution of student attitudes throughout their studies may be instructive in identifying those at the highest risk of disengaging and/or discontinuing. Certain HE experiences may drive down student motivation. For example, if students perceive that their **course misaligns with their occupational aspirations**, they may be less inclined to continue their studies and thus present a higher discontinuation risk.[15] Insight into whether the programme of study was a student's first, second, or third choice may provide some inference into motivation levels, though one study researching this area did not find that it impacted discontinuation rates.[16] Furthermore, a low or **lack of confidence in the subject matter** may negatively affect student engagement.[17] Such patterns are not restricted to in-person or hybrid learning environments, with research showing that online education is analogously affected.[18] UoW collects data that could be used as proxies for these indicators. For instance, the Student Module Evaluation (SME) asks students whether 'The module is intellectually stimulating'. A negative response to this question may indicate a low level of motivation.

Aside from course or academic stimuli, experiences in HE may negatively impact student **wellbeing** and, by extension, attitudes towards study. Northumbria University found that collecting, monitoring and integrating student wellbeing data with the much larger dataset they analysed using their predictive model enabled them to identify more at-risk students than educational analysis alone. They also ascertained that it was possible 'to predict a student's wellbeing with significant accuracy'. The wellbeing data at the base of this approach was derived from the World Health Organisation's Five Well-Being Index (WHO-5) (Appendix 1), which students had the option of completing initially during enrolment and at set points throughout the year.[19] Furthermore, Crane *et al.*'s study reported that the experience of microaggressions had a limiting effect on the academic persistence of LGBTQ (Lesbian, Gay, Bisexual, Transgender,

[15] Steffen Wild, "Trajectories of subject-interests development and influence factors in higher education," *Current Psychology* 42 (2022).
[16] Liga Paura and Irina Arhipova, "Cause Analysis of students' dropout rate in higher education study program," *Procedia – Social and Behavioural Sciences* 109 (2014).
[17] Kevin Paterson and Adam Guerrero, "Predictive Analytics in Education: Considerations in Predicting versus Explaining College Student Retention," *Research in Higher Education Journal* 44 (2023): 1-12.
[18] Olga Rotar, "A missing theoretical element of online higher education student attrition, retention, and progress: a systematic literature review," *SN Social Sciences* 2, no.12 (2022).
[19] Jim Keane, *The Office for Students (OfS) mental health analytics project: an evaluation*, Jisc, 2024, 2 and 6.

Queer) students. Microaggressions contributed to greater intentions of LGBTQ students to transfer from their university. In other words, as the experience of microaggressions increased, classroom discomfort also increased, which lessened the intention to persist.[20] Relatedly, institutional research found that a lack of a sense of belonging can precipitate disengagement and potentially discontinuation.[21] Therefore, the research suggests that to aid in the identification of at-risk students, study **motivation**, **sense of belonging**, and **wellbeing** need to be monitored; this requires identifying the proxy data already collected by the institution or its creation. For instance, the annual Transformation in Students Survey (TiSS) asks students for their level of (dis)agreement with the following statements: 'I feel I have a deep sense of belonging at university' and 'I have developed coping strategies for dealing with stressful situations at university'.

In addition to the range of variables that can relate to student attitudes, those of students' wider networks can also shape decisions concerning (non-)continuation. Several studies have highlighted the role of perceived and actual **family support**.[22] Family support can take various forms, including financial and emotional, which are crucial for student retention. The converse of this is that family instability emerges as a risk factor.[23] While identifying a suitable proxy as an indicator of a "supportive family environment" might be necessary, it may also be beneficial to examine any continuation risks associated with **care-experienced** and **estranged students**.

Model builders and HE providers have demonstrated a propensity to include various academic variables in modelling predictive LA.[24] Alongside behavioural factors, academic variables are some of the strongest predictors of disengagement and discontinuation. Academic grades of numerous types have been utilised in LA and predictive LA models.[25] Wild identified a 'descending interest trajectory'. In other words, there is a group of students who, over time, experience a decline in subject interest. This may be due to the attitudinal factors discussed above, but it may also be related to **relatively low university entrance scores**.[26] Indeed, scholars

---

[20] Phoenix R. Crane, Katarina S. Swaringen, Matthew M. Rivas-Koehl, Anthony M. Foster, Tran H. Le, Dana A. Weiser, and Amelia E. Talley, "Come Out, Get Out: Relations Among Sexual Minority Identification, Microaggressions, and Retention in Higher Education," *Journal of Interpersonal Violence* 37, no. 9-10 (2022).

[21] Rawlinson and Alexander, "Perspectives on Student (Dis)Engagement and Continuation."

[22] Tasos Stylianou and Alexandros Milidis, "The socioeconomic determinants of University dropouts: The case of Greece," *Journal of Infrastructure, Policy and Development* 8, no.6 (2024); Michael Jüttler, "Predicting economics student retention in higher education: The effects of students' economic competencies at the end of upper secondary school on their intention to leave their studies in economics," *PLoS ONE* 15, no.2 (2020): 1-27; Nurmalitasari, Zalizah Awang Long and Mohammad Faizuddin Mohd Noor, "Factors Influencing Dropout Students in Higher Education," *Education Research International* (2023); and Valentim Realinho, Jorge Machado, Luis Baptista and Monica V. Martins, "Predicting Student Dropout and Academic Success," *Data* 7, no.146 (2022).

[23] Archambault *et al*. "Student Engagement and School Dropout."

[24] Sghir, Adadi and Lahmer, "Recent advances in Predictive Learning Analytics."

[25] Oliveira *et al*., "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education."

[26] Wild, "Trajectories of subject-interests development and influence factors in higher education." See also Yasmin Erika Faridhan, Birgit Loch and Lyndon Walker, "Improving retention in first-year mathematics using learning analytics," in: Electric Dreams, 30th ASCILITE Conference, Macquarie University, Sydney, 1-4 December 2013.

have found **pre-HE academic performance** to be a valuable indicator of discontinuation, with those with lower grades being at higher risk of ceasing their studies.[27] Institutional research has identified that language barriers for international students may precipitate disengagement. Thus, **scores from language assessment tests** (i.e., International English Language Testing System) should be investigated to ascertain if there is a link to retention rates.[28] Higher **assessment** and **final scores** at HE have been found to be correlated with higher student engagement.[29] In US studies, low **grade point averages** (GPAs) have been identified as significant predictors of discontinuation.[30] Freitas found academic performance to be strongly related to graduation rates, with students receiving average marks in the 60s or 70s being more likely to graduate.[31] Similar outcomes have been found for in-person or hybrid and online students.[32] At a more granular level, research has shown that **in-class performance on tests**, etc., correlated with higher engagement.

There is also some correlation between **study load** (i.e., part-time, full-time, online, hybrid) and retention. For example, Jia and Maloney found that students on **part-time** programmes were more vulnerable to non-completion; indeed, being a part-time student increased this probability by 12.231 percentage points. The same study found that part-time students were 'substantially more likely to drop out of university in the second year':

> 'Studying part-time increases the probability of non-retention by an average of 18.037 percentage points. Thus, part-time study is arguably the single most important single at risk factor for poor university outcomes.'[33]

---

[27] Paura and Arhipova, "Cause Analysis of students' dropout rate in higher education study program"; Li *et al.*, "Retention Factors in STEM Education Identified Using Learning Analytics"; Faridhan, Loch and Walker, "Improving retention in first-year mathematics using learning analytics"; Jüttler, "Predicting economics student retention in higher education." See also Pengfei Jia and Tim Maloney, "Using predictive modelling to identify students at risk of poor university outcomes," *Higher Education* 70 (2015); and Umair Uddin Shaikh and Zaheeruddin Asif, "Persistence and Dropout in Higher Online Education: Review and Categorisation of Factors," *Frontiers in Psychology* (2022); Ghaith Al-Tameemi, James Xue, Suraj Ajit, Triantafyllos Kanakis, Israa Hadi, "Predictive Learning Analytics in Higher Education: Factors, Methods and Challenges," in: 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, June 2020): 1-9; Sandra C. Matz, Christina S. Bukow, Heinrich Peters, Christine Deacons, Alice Dino and Clemens Stachl, "Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics," *Nature: Scientific Report* 13, 5705 (2023): 1-16; Aina *et al.*, "The determinants of university dropout."

[28] Scott Rawlinson, "Academic and Practical Information Seeking Behaviours and Needs of International Students at Pre-arrival and Arrived (First Year) Stages," (University of Westminster, 2023); Al-Tameemi *et al.*, "Predictive Learning Analytics in Higher Education."

[29] Annika Wolff, Zdenek Zdrahal, Andriy Nikolov, and Michal Pantucek, "Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment," in: *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, Association for Computing Machinery, New York, NY, USA, pp. 145–149, 2013; Al-Tameemi *et al.*, "Predictive Learning Analytics in Higher Education."

[30] *Ibid.*; Paterson and Guerrero, "Predictive Analytics in Education"; Lovenoor Aulck, Nishant Velagapudi, Joshua Blumenstock and Jevin West, "Predicting Student Dropout in Higher Education," in: 2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications; Nurmalitasari, Long and Noor, "Factors Influencing Dropout Students in Higher Education"; Realinho, "Predicting Student Dropout and Academic Success."

[31] Sara de Freitas, David Gibson, Coert Du Plessis, Pat Halloran, Ed Williams, Matt Ambrose, Ian Dunwell and Sylvester Arnab, "Foundation of dynamic learning analytics: Using university student data to increase retention," *British Journal of Educational Technology* 46, no.6 (2015): 1175-1188.

[32] Justin D. Cochran, Stacy M. Campbell, Hope M. Baker and Elke M. Leeds, "The Role of Student Characteristics in Predicting Retention in Online Courses," *Research in Higher Education* 55 (2014): 27-48.

[33] Jia and Maloney, "Using predictive modelling to identify students at risk of poor university outcomes," 141 and 144.

Rodríguez-Muñiz *et al*. found that part-time students were at higher risk of discontinuation than their full-time peers.[34] Regarding **online and blended courses**, the driving force underscoring Rotar's study was a desire to understand high discontinuation rates for online learners.[35] Furthermore, Fan *et al*. found that students taking online or blended courses were not equipped with adequate skills or support, leading to associated issues of disengagement and attrition; however, generalisability may be limited due to the research being confined to one Australian HE institution.[36] Additionally, **admission status**, particularly being admitted through non-standard entry pathways (i.e., special admissions, internal bridging programs), increased the risk of non-retention.[37]

Evidence from the literature suggests that the modelling of student attrition should not be restricted to students' pre-HE and HE academic performance. Indeed, predictive models should also consider the **highest parental level of education**, with students from families with higher educational attainment being more likely to persist in their studies.[38]

3.2 Student Demographics

Numerous studies noted how **student demographics** have been fed into predictive LA models.[39] Including student characteristics in predictive LA models has been shown to improve their prognostic capacity regarding retention.[40] The efficacy of their inclusion has been demonstrated when combined with student engagement and behavioural data.[41] This comes with the caveat that the most significant demographic variables will vary from institution to institution.[42] Nevertheless, getting an overview of factors typically included in predictive LA models is helpful. The picture concerning **sex/gender** is a mixed one, highlighting the importance of institutional context. Some studies have found female students to be at higher risk of disengagement than their male peers.[43] Faridhan, Loch and Walker found issues with female retention in STEM

[34] Luis J. Rodríguez-Muñiz, Ana B. Bernardo, María Esteban and Irene Díaz, "Dropout and transfer paths: What are the risky profiles when analyzing university persistence with machine learning techniques?" *PLOS ONE* 14, NO.6 (2019), 15.
[35] Rotar, "A missing theoretical element of online higher education student attrition, retention, and progress."
[36] Si Fan, Allison Trimble, David Kember, Tracey Muir, Tracy Douglas, Yanjun Wang, Jennifer Masters and Casey Mainsbridge, "Supporting engagement and retention of online and blended-learning students: A qualitative study from an Australian University," *The Australian Education Researcher* 11 (2023): 1-19.
[37] Jia and Maloney, "Using predictive modelling to identify students at risk of poor university outcomes."
[38] Dawson *et al*., "From prediction to impact"; Stylianou and Milidis, "The socioeconomic determinants of University dropouts"; Aina *et al*., "The determinants of university dropout."
[39] Oliveira *et al*., "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education"; Sghir, Adadi and Lahmer, "Recent advances in Predictive Learning Analytics".
[40] Wolff *et al*., "Improving retention"; Al-Tameemi *et al*., "Predictive Learning Analytics in Higher Education."
[41] Matz *et al*., "Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics."
[42] Wolff *et al*., "Improving retention."
[43] Wild, "Trajectories of subject-interests development and influence factors in higher education"; Cochran *et al*., "The Role of Student Characteristics in Predicting Retention in Online Courses"; Aina *et al*., "The determinants of university dropout"; Realinho, "Predicting Student Dropout and Academic Success."

(Science, technology, engineering, and mathematics) subjects.[44] However, Jia and Maloney found that being female lowered the probability of non-completion.[45] In a quantitative study analysing data from Latvian HE, Paura and Arhipova found that male students had a 1.5 times higher risk of discontinuation than female students.[46] Other studies have identified that male and female students leave their studies for different reasons, with females more likely to attribute discontinuation to poor academic performance or wrong faculty choice. In contrast, males were more affected by failing a course.[47] This diversity highlights the importance of examining institutional data and reinforces the necessity of including sex/gender as a variable in predictive modelling.[48]

A similar case can be made regarding a multitude of demographic data points. For instance, given that it is known to have an impact on retention, the **disability status** of students has been fed into predictive models.[49] Disparities in non-completion among students of different ethnicities affirm the importance of factoring in **ethnicity** as a variable in predictive modelling.[50] Additionally, **socio-economic** or class background has been shown to affect retention rates. For instance, students from lower decile schools, which represent poor socio-economic backgrounds, face a greater risk of non-completion.[51] **Age** has also been shown to be a helpful predictor of non-completion; however, varying definitions of "older" or "mature" students make it difficult to talk in generalities.[52] As a demonstration of this diversity, Jia and Maloney have argued that older students, especially those over 30, show a higher risk of non-completion and non-retention than younger students, while elsewhere, that figure was 24 and 23.[53] Elsewhere,

---

[44] Faridhan, Loch and Walker, "Improving retention in first-year mathematics using learning analytics." See also Gabriela Ortiz-Martínez, Patricia Vázquez-Villegas, María Ileana Ruiz-Cantisani, Mónica Delgado-Fabián, Danna A. Conejo-Márquez and Jorge Membrillo-Hernández, "Analysis of the retention of women in higher education STEM programs," *Humanities and Social Sciences Communications* 10, no.101 (2023): 1-14.
[45] Jia and Maloney, "Using predictive modelling to identify students at risk of poor university outcomes," 141. See also Carmen Aina, Eliana Baici, Giorgia Casalone and Francesco Pastore, "The determinants of university dropout: A review of the socio-economic literature," *Socio-Economic Planning Sciences* 79 (2022): 1-16.
[46] Paura and Arhipova, "Cause Analysis of students' dropout rate in higher education study program."
[47] Stylianou and Milidis, "The socioeconomic determinants of University dropouts."
[48] Shaikh and Asif, "Persistence and Dropout in Higher Online Education."
[49] Carol Elaine Calvert, "Developing a model and applications for probabilities of student success: a case study of predictive analytics," *Open Learning: The Journal of Open, Distance and e-Learning* 29, no.2 (2014): 160-173; Wolff *et al*., "Improving retention."
[50] Jia and Maloney, "Using predictive modelling to identify students at risk of poor university outcomes"; Kameryn Denaro, Kimberly Dennin, Michael Dennin and Brian Sate, "Identifying systemic inequity in higher education and opportunities for improvement," *PLoS ONE* 17, no.4 (2022): 1-16; Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics."
[51] Jia and Maloney, "Using predictive modelling to identify students at risk of poor university outcomes"; Archambault *et al*. "Student Engagement and School Dropout"; Calvert, "Developing a model and applications for probabilities of student success"; Faridhan, Loch and Walker, "Improving retention in first-year mathematics using learning analytics"; Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics." See also: Paterson and Guerrero, "Predictive Analytics in Education"; Wolff *et al*., "Improving retention"; Aina *et al*., "The determinants of university dropout"; Realinho, "Predicting Student Dropout and Academic Success."
[52] Faridhan, Loch and Walker, "Improving retention in first-year mathematics using learning analytics"; Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics"; Rotar, "A missing theoretical element of online higher education student attrition, retention, and progress"; Shaikh and Asif, "Persistence and Dropout in Higher Online Education"; Aina *et al*., "The determinants of university dropout"; Realinho, "Predicting Student Dropout and Academic Success."
[53] Jia and Maloney, "Using predictive modelling to identify students at risk of poor university outcomes," 141; Cochran *et al*., "The Role of Student Characteristics in Predicting Retention in Online Courses"; Rodríguez-Muñiz *et al*., "Dropout and transfer paths."

Freitas found that students closer to the average age of their cohort had higher retention rates.[54] Another possible dimension concerns **whether students receive financial support** and **the nature of that support**.[55] In one study, it was found that students in receipt of loans were more likely to withdraw than those receiving merit-based scholarships.[56] **Study level** (i.e., undergraduate, postgraduate taught, postgraduate research) is another important variable to factor into modelling.[57] As with all these variables, it is essential to investigate and understand how they express themselves within an institutional context and interact with attitudinal, academic, and behavioural factors.

The variables mentioned thus far do not represent the outermost limit of factors potentially shaping student decision-making. For example, **students who travel a significant distance** to their university/campus (i.e., commuter students) may be at greater risk of discontinuation, with studies finding a negative association between distance and performance and retention.[58] One study discovered that distance from campus had a detrimental impact on graduation rates, with rates falling for students living more than 30km away.[59] Beyond proximity to campus, **family responsibilities** such as childcare were found to impact persistence in online learning.[60] While not discussed in this study, the impact of parenting responsibilities on continuation raises the issue of caring responsibilities more generally and the inclusion of **caring status** in predictive modelling.[61]

3.3 Student Behaviours

Different types of engagement with VLE/LMS platforms have featured prominently in the literature.[62] Indeed, studies have noted strong correlations between changes in students' VLE activity and their likelihood of discontinuation.[63] For example, Wolff found that sudden drops in VLE engagement just before assessments signalled a higher likelihood of course failure and discontinuation.[64] At a more granular level, the types of **VLE engagement** tracked covered a range of activities, including interaction with the VLE homepage, course content, **chatting** (i.e., student

[54] Freitas *et al*., "Foundation of dynamic learning analytics."
[55] Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics."
[56] Cochran *et al*., "The Role of Student Characteristics in Predicting Retention in Online Courses."
[57] *Ibid*.
[58] Dawson *et al*., "From prediction to impact."
[59] Freitas *et al*., "Foundation of dynamic learning analytics."
[60] Shaikh and Asif, "Persistence and Dropout in Higher Online Education."
[61] Scott Rawlinson, "'Little Islands': challenges and opportunities for student carers in higher education," *International Journal of Inclusive Education* (2024): 1-16.
[62] Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics"; Edward Peck, *Student analytics: A core specification for engagement and wellbeing analytics* (Jisc, 2023), 6.
[63] Beer *at el*., "Measuring engagement."
[64] Wolff *et al*., "Improving retention."

interactions), **completion of assessment activities** and forum activities, number of **logins**, and overall **time spent** on the VLE.[65]

One study commented that engagement metrics could be used as a proxy for continuation. It refers to 'core specification data' that had demonstrated a correlation between engagement and continuation. Predictably, **attendance at scheduled teaching sessions** is featured in this list, echoing other research.[66] Relatedly, Freitas *et al*. found that greater use of online materials and on-site attendance decreased attrition, with the use of Blackboard and library resources being associated with higher graduation rates.[67] However, it is important to acknowledge that there is a diversity of behaviours as far as engagement with online learning environments is concerned, and a student with minimal engagement may still pass. Furthermore, Peck stated that '**library usage**, such as taking out book loans' was a useful metric to ascertain engagement and identify any signs of disengagement.[68]

Other student behaviours can provide some instruction on the likelihood of a student becoming disengaged or quitting their course. Research has found that students who have **previously withdrawn** from their studies are more likely than their peers to withdraw again.[69]

3.4 Institutional Factors: Education, Teaching and the University Community

Several institutional factors have been shown to have a significant bearing on students' decisions to (dis)continue their studies. These include **perceptions of teaching quality**, the **quality and relevance of course content**, course design, and class size.[70] Studies have suggested that poor quality teaching can exacerbate the decline of student interest in their course or programme.[71] A wealth of information on teaching quality can be gleaned from surveys already conducted at Westminster. For example, the National Student Survey (NSS) could be mined for historical data on satisfaction, while SMEs could provide a closer approximation of real-time information; the latter includes a question on whether students felt 'Teaching staff are good at explaining things'

---

[65] Al-Tameemi *et al*., "Predictive Learning Analytics in Higher Education"; Dawson *et al*., "From prediction to impact;" Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics"; Ewa Seidel and Salah Kutieleh, "Using predictive analytics to target and improve first year student attrition," *Australian Journal of Education* 61, no.2 (2017): 200-218; Raj and Renumol, "Early prediction of student engagement in virtual learning environments using machine learning techniques"; Peck, *Student analytics*, 6.

[66] Peck, *Student analytics*, 6; Rawlinson and Alexander, "Perspectives on Student (Dis)Engagement and Continuation"; Al-Tameemi *et al*., "Predictive Learning Analytics in Higher Education"; Freitas *et al*., "Foundation of dynamic learning analytics;" Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics"; Faridhan, Loch and Walker, "Improving retention in first-year mathematics using learning analytics"; Oliveira *et al*., "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education."

[67] Freitas *et al*., "Foundation of dynamic learning analytics."

[68] Peck, *Student analytics*, 6.

[69] Cochran *et al*., "The Role of Student Characteristics in Predicting Retention in Online Courses."

[70] Li *et al*., "Retention Factors in STEM Education Identified Using Learning Analytics"; Rawlinson and Alexander, "Perspectives on Student (Dis)Engagement and Continuation"; Nurmalitasari, Long and Noor, "Factors Influencing Dropout Students in Higher Education."

[71] Wild, "Trajectories of subject-interests development and influence factors in higher education."

as well as the statement ‘Overall, I am satisfied with this module’. This has the advantage of removing the need to create new data points. Instead, integrating SME and other survey data with information on retention should be a priority. In terms of course and class size, Jia and Maloney identified a link between larger **class sizes** and **high student-to-teacher ratios** and increased likelihood of non-completion.[72] **Course content and design** have been identified as a predictor variable for student attrition, with models considering variables such as number of assignments, course pass rates, and course content relevance (i.e., alignment with career or personal development goals).[73]

3.5 Indicators of Student Disengagement and Discontinuation

The REA enabled the tabulation of (dis)engagement and (dis)continuation indicators (Figure 3). In terms of adopting a predictive model, it is recommended that real-time/historical data covering the indicators listed below constitute its foundations.

Figure 3. Indicators of Student Disengagement and Discontinuation

No.	Indicator	Description and Evidence	Data source
Student Demographic Indicators			
1	Admission status	Entry via non-standard routes (i.e., special admissions and internal bridging programs) has been linked with an increased risk of non-retention [22].	SITS (existing)
2	Age	Age has been identified as a useful predictor of non-completion, with “older” students at particular risk [1, 9, 15, 17, 22, 27, 42, 43, 44, 49].	SITS (existing)
3	Care-experienced status	This term refers to students who have lived in care at any stage of their life. Studies have identified that family support (e.g., financial and/or emotional) is crucial for student retention. Conversely, family instability represents a risk factor [23, 51].	SITS (existing)
4	Caring responsibility status	This term encompasses “traditional” carers who look after ill relatives or friends, as well as student parents/guardians [41, 49].	SITS (existing)
5	Disability status	This metric covers physical and mental disability. Research has identified that continuation rates for disabled students are lower than for peers with no reported disability [8, 19, 56].	SITS (existing)
6	Distance from primary campus	Students who travel a significant distance to university/campus (i.e., commuter students) have been shown, in some cases, to be at increased risk of discontinuation [12, 17].	SITS (existing)
7	English as first language	Is English a student’s first language? How did students perform in English-language competency testing? Research suggests that students entering UK HE with relatively lower English-language competence scores are at higher risk of disengagement [25].	SITS (existing)
8	Estranged student status	This term refers to any young person under 25 classified as an independent student on the grounds of estrangement by	SITS (existing)

[72] Jia and Maloney, “Using predictive modelling to identify students at risk of poor university outcomes”; Aina *et al.*, “The determinants of university dropout.”

[73] Li *et al.*, “Retention Factors in STEM Education Identified Using Learning Analytics”; Sghir, Adadi and Lahmer, “Recent advances in Predictive Learning Analytics”; Rotar, “A missing theoretical element of online higher education student attrition, retention, and progress.”

		Student Finance England. Studies have identified that family support (e.g., financial and/or emotional) is crucial for student retention. Conversely, family instability represents a risk factor [23, 51].	
9	Ethnicity	Numerous studies have found ethnic disparities in non-completion among students [13, 22, 27].	SITS (existing)
10	Family support	Studies have identified that family support (e.g., financial and/or emotional) is crucial for student retention. Conversely, family instability represents a risk factor [23, 30, 42, 51].	
11	Financial support	Whether and what financial support a student receives has been correlated with (dis)engagement [9, 27, 33].	SITS and Student Funding Team (SFT) (existing)
12	First-in-family	Whether or not a student was the first in their family to attend university was ranked among the top 20 predictors of nearly 800 variables in Northumbria University's predictive model [27, 36]	SITS (existing)
13	Highest level of parental education	This metric refers to the highest level of education attained by any parent residing in the same household as the child/young adult. Research has shown that students from families with higher educational attainment are more likely to persist in their studies [12, 27, 51].	SITS (existing)
14	Household income	Household income is a significant factor in student retention. Students from lower-income households are more likely to withdraw [27].	SFE
15	IMD quintile	A measure of relative deprivation linked to disengagement, students from lower quintiles are at greater risk [19].	SITS (existing)
16	Marital status	Legally defined marital status: single, married, widowed, divorced. [30, 42]	SITS
17	Parental occupation	Parental occupation has been linked to disengagement. A student's socio-economic status, household income, or IMD quintile could act as a proxy for parental occupation [1, 27].	SITS/SFE
18	Sex/gender	The impact of sex/gender on (non-)continuation has been shown to vary from institution to institution, with some research indicating females are at greater risk and others that men are [1, 9, 15, 22, 32, 35, 42, 49, 51, 55].	SITS (existing)
19	Socio-economic or class background	Socio-economic status (SES) may be gleaned from IMD (Index of Multiple Deprivation) quintile data or information on household income from Student Finance England. Research has shown that SES affects retention rates [1, 3, 8, 15, 22, 27, 32, 42, 56].	SITS (existing)
20	Study level	Study level (i.e., foundation, undergraduate, postgraduate taught, and postgraduate research) has been linked to retention [9]	SITS (existing)
21	Study load	Refers to full-time or part-time study commitment. Studies have found a correlation between part-time study and vulnerability to non-completion [22, 43].	SITS (existing)
22	Study mode	Study mode refers to whether courses are in-person, hybrid or online. Several pieces of research have noted the high discontinuation rates for online learners [14, 44].	SITS (existing)
23	Types of school	England hosts a variety of school types, including faith schools, free schools, academies, city technology colleges, state boarding schools, and private schools. The type of secondary school has been observed as a contributory factor regarding academic performance and retention [2, 12]	SITS (existing)
24	Working whilst studying	Does the student work a job whilst studying [19, 27, 30]?	SITS (existing)
Academic Indicators			
25	In-class, module and end-of-year assessment scores	Assessment scores are an important indicator of retention [2, 6, 9, 17, 27, 30, 34, 42, 56].	SITS (existing)

26	English language proficiency assessment score	Research indicates that language barriers for international students may precipitate disengagement [2, 39].	SITS (existing)
27	Pre-HE academic performance	Students with lower HE entry grades have been identified as at higher risk of ceasing their studies [1, 2, 15, 22, 23, 27, 29, 35, 38, 49]	SITS (existing)
28	Programme of study rank	Was the student's programme of study their first, second, or third choice [35]?	
29	University entrance scores	Low university entrance scores have been linked to disengagement [15, 55]	SITS (existing)
Engagement and Behavioural Indicators			
30	Attendance at scheduled teaching and other sessions	Do students regularly attend scheduled teaching, seminars, and other sessions, and are there any gaps in attendance [2, 15, 17, 31, 36, 53]?	SITS (existing)
31	Library usage	Are students accessing the library or taking out book loans [36, 53]?	Library engagement data
32	Mitigating circumstances claims, including if it had been rejected	Students whose studies have been disrupted due to unforeseen circumstances or incidents outside their control may submit a mitigating circumstances claim. Studies have found a correlation between submissions of claims and attrition. NB. models should take into account successful and unsuccessful claims [25].	
33	Motivation level	Research has found that certain experiences can drive down student motivation, such as when course content misaligns with students' occupational aspirations [30, 55] or lack confidence in the subject matter [34, 44], or whether the degree being undertaken was the student's first choice [27].	
34	(Non-)payment of tuition fees	Are students up to date with the payment of their tuition fees [56]?	Finance data
35	Sense of belonging	Lack of belonging has been demonstrated to precipitate disengagement and potentially discontinuation [11, 40].	TiSS
36	VLE engagement – chatting	Do students communicate with other students and lecturers during and outside sessions using VLE messaging functionality? Are there any gaps in communication? [2, 7, 12, 17, 27, 36, 38, 46, 56]?	Blackboard
37	VLE engagement – clicks (total) on VLE	The total number of clicks/interactions with the VLE [7, 38].	Blackboard
38	VLE engagement – completion of assessment and forum activities	Are students regularly completing any assessment activities on the VLE? Are there any gaps in assessment or activity completion? How many words or question marks are used in forum posts [2, 7, 12, 17, 27, 36, 38, 46, 56]?	Blackboard
39	VLE engagement – course content	How frequently do students engage with course content on the VLE? This could be the number of lectures or videos viewed [2, 7, 12, 17, 27, 36, 38, 46, 56].	Blackboard
40	VLE engagement – first day of access	The timing of each student's initial access of the VLE each term [7, 27].	Blackboard
41	VLE engagement – number of logins	How frequently do students log in to the VLE [2, 7, 12, 17, 27, 36, 38, 46, 53, 56]?	Blackboard
42	VLE engagement – time spent (total)	How frequently and for what duration of time do students engage with the VLE [2, 7, 12, 17, 27, 36, 38, 46, 56]?	Blackboard
43	Withdrawal history	Does the student have a history of withdrawal? Research has shown that students who have previously withdrawn from their studies are more likely than those with no history of withdrawal [9].	SITS
Institutional Indicators			
44	Cohort/class size	What is the size of the cohort/class on the course, and what is the colleague-student ratio [1, 22]?	QlikView
45	Course design	Course design refers to its content, structure, and learning outcomes. Are students (dis)satisfied with their course design [27, 44, 47]?	NSS

46	Relevance of course content	Does the course content match students' expectations and aspirations (i.e., career or personal development goals) [27, 40]?	NSS
47	Teaching quality	What do students think about the teaching quality in their module and/or course [13, 27, 30, 40, 55]?	SME, NSS
Wellbeing Indicators			
48	Responses to WHO-5 well-being questionnaire	The WHO-5 comprises five questions that measure an individual's wellbeing. The questionnaire has clinical validity for assessing wellbeing [25, 36]	WHO-5 (New)

3.6 Conclusion

It is helpful to think of discontinuation as a recipe; rarely is a single or isolated demographic or academic factor sufficient to predict the cessation of studies. Rather, combining multiple ingredients helps account for a student's decision not to complete their studies. For example, Ortiz-Martínez *et al.* found retention issues among female students pursuing STEM subjects.[74] Denaro *et al*. explored the intersection of ethnicity/race and course programmes (specifically STEM). They found that ethnic minority students performed less well academically, felt lower levels of belonging, and had lower graduation rates than their peers.[75] Cochran noted an intersection between the type of financial support and the persistence of Black students in online learning. Thus, Black students with merit-based scholarships were found to be less likely to withdraw, but those without loans were at greater risk of withdrawal. This study also identified interconnections between programme choice, sex, ethnicity and academic performance.[76] Additionally, the pressures of family responsibilities and parenting have been found to impact women disproportionately. The effects of demographics and attitudinal factors are intricately tied to the institutional setting. For instance, the discomfort felt by LGBTQ students was enabled by a wider institutional context.[77] Additionally, while a lack of a sense of belonging can fuel disengaging behaviours, HE institutions can nurture feelings of belonging via initiatives such as "Decolonising the Curriculum".[78]

4.0 Designing LA and Predictive LA Models

Multiple and complexly interrelated factors shape student engagement, strengthening the case for predictive LA models, which can handle the abundance of behavioural, demographic, academic, and psychological attributes and establish patterns indicating which students are at the highest risk.[79] HE providers and others have developed a range of predictive models to tackle

[74] Ortiz-Martínez *et al*., "Analysis of the retention of women in higher education STEM programs."
[75] Denaro et al., "Identifying systemic inequity in higher education and opportunities for improvement."
[76] Cochran *et al*., "The Role of Student Characteristics in Predicting Retention in Online Courses."
[77] Crane *et al*., "Come Out, Get Out."
[78] Rawlinson and Alexander, "Perspectives on Student (Dis)Engagement and Continuation."
[79] Kahu, "Framing student engagement in higher education"; Oliveira *et al*., "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education"; Shafiq *et al*., "Student retention using educational data mining and predictive analytics."

student attrition using data mining and ML techniques.[80] These have included supervised learning approaches, such as decision trees, as well as logistic regression and artificial neural networks (ANN); there has been a relative dearth of unsupervised and hybrid approaches.[81] Several studies opted for logistic regression and/or discriminant analysis.[82] Logistic regression is commonly used when the dependent variable is binary (i.e., continue = 0, discontinue = 1). For instance, Paterson and Guerrero compared the performance and accuracy of logistic regression and discriminant analysis to predict the risk of student discontinuation. Their model comprised numerous variables, including high-school GPA and American College Testing (ACT) score, among other academic performance indicators. Their study found that discriminant analysis (85.76% predictive accuracy) outperformed logistic regression (84.09% predictive accuracy).[83] As such models have matured with the analysis of more data, improvements in their accuracy have been observed.[84] ANNs are a 'commonly used' type of ML model designed to learn patterns from data. Like the brain, ANNs consist of a web of interconnected neurons that each process a small piece of information and pass it on to the next neuron. Through training, these models can learn to recognise patterns in data.[85]

Random Forest is also a popular ML technique adopted by HE providers. Indeed, Northumbria University's model was based on this technique. A Random Forest is an ML algorithm that combines multiple decision trees (i.e., a diagram showing the possible outcomes of a series of decisions) to make predictions.[86] Or, as Keane explains, 'This approach uses a subset of the source data to build many multiple random decision trees, which are iterated multiple times until the model reaches a degree of accuracy in predicting the outcome for the training data. It is then tested on data it hasn't seen to assess its accuracy.'[87] The Northumbria University model contains over 800 static and dynamic variables derived from an REA, interviews with wellbeing staff and a case management review.[88] Matz *et al*. assessed the predictive performance of Random Forest concerning student attrition and concluded that the technique had higher predictive accuracy than the elastic net model and that models using institutional and

---

[80] Oliveira *et al*., "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education."
[81] Shafiq *et al*., "Student retention using educational data mining and predictive analytics." Freitas *et al*., "Foundation of dynamic learning analytics," detailed a semi-supervised neural network algorithm.
[82] Aulck *et al*., "Predicting Student Dropout in Higher Education"; Paterson and Guerrero, "Predictive Analytics in Education"; Calvert, "Developing a model and applications for probabilities of student success"; Herodotou *et al*., "How Can Predictive Learning Analytics and Motivational Interventions Increase Student Retention and Enhance Administrative Support in Distance Learning?"
[83] Paterson and Guerrero, "Predictive Analytics in Education."
[84] Calvert, "Developing a model and applications for probabilities of student success."
[85] Al-Tameemi *et al*., "Predictive Learning Analytics in Higher Education."
[86] Niklas Donges, "Random Forest: A Complete Guide for Machine Learning," *Builtin*. [Accessed 16/01/2025]. https://builtin.com/data-science/random-forest-algorithm.
[87] Keane, *The Office for Students (OfS) mental health analytics project*.
[88] Office for Students, "Mental health analytics: An innovative approach to understanding students' wellbeing." *Office for Students* [Accessed 17/01/2025]. https://www.officeforstudents.org.uk/for-providers/equality-of-opportunity/effective-practice/mental-health-analytics-an-innovative-approach-to-understanding-students-wellbeing/.

engagement data outperformed those using only one type of data.[89] Additionally, Raj and Renumol tested various ML models, including decision trees, logistic regression, Random Forest, K-Nearest Neighbours and evaluated them for accuracy, precision and recall. Random Forest was the most effective algorithm, achieving 94.1% accuracy, 94.9% precision and 97.4% recall in identifying at-risk students early in a course.[90]

Integrating predictive LA models into existing institutional data systems is crucial for its effectiveness. As such, UoW must prioritise its technical readiness to implement predictive LA by ensuring seamless data integration. Institutional data from disparate systems and datasets, including enrolment, academic performance, and student interactions with student support systems, needs to be integrated to create a holistic view of the student so that their risk of discontinuation can be calculated with a high degree of accuracy. For example, internal data sharing is at the heart of Northumbria University's approach as 'being able to see that a student is not attending classes, showing declining grades, not paying their tuition fees, accessing learning resources in particular ways and not leaving their accommodation suggests a student in need of support.'[91]

Besides technical readiness, adopting predictive LA entails serious consideration of several potential pitfalls and challenges. These cover data quality, protection, governance, ethics, as well as faculty and leadership buy-in. Predictive LA models are built on vast foundations of student data, which has led to concerns about privacy and ethics.[92] Effective governance and accountability are required to ensure that the usage of predictive LA does not become intrusive.[93] Students need to be clear on the purposes for which their data is being used, and informed consent should be sought. Northumbria University opted for this lawful basis as although it risks excluding students who opt out of proactive monitoring, data processing is legal, and students are informed about how their data is being used.[94] Other ethical concerns include the potential for miscalculation of students' risk levels in a way that might lead to unintended consequences such as demotivation or complacency.[95] In addition to ensuring the technical infrastructure is in place, Huijser, West and Heath have argued that ongoing buy-in from senior

[89] Matz *et al.*, "Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics."
[90] Raj and Renumol. "Early prediction of student engagement in virtual learning environments using machine learning techniques."
[91] Keane, *The Office for Students (OfS) mental health analytics project*.
[92] Oliveira *et al.*, "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education"; Shafiq *et al.*, "Student retention using educational data mining and predictive analytics."
[93] Guzmán-Valenzuela *et al.*, "Learning analytics in higher education."
[94] Keane, *The Office for Students (OfS) mental health analytics project*, 7.
[95] Teo Susnjak, Gomathy Suganya Ramaswami and Anuradha Mathrani, "Learning analytics dashboard: a tool for providing actionable insights to learner," *International Journal of Educational Technology in Higher Education* 19, no.12 (2022): 1-23.

leadership and alignment with strategic priorities is essential.[96] Senior leadership buy-in must be sustained, and integration of predictive LA with retention goals and institutional strategies may be one method of ensuring a constancy in interest. Additionally, such tools are likely to have pertinence for a wide range of staff working in different areas of the university. Conducting stakeholder analysis and maintaining ongoing collaboration with academics, IT staff, administrative leaders, etc., should be a top consideration.[97]

Furthermore, the literature indicates that levels of data literacy among academics and other university staff may vary.[98] Low levels of data literacy may have the unwanted effect of causing frustration, indifference, or even opposition to predictive LA tools. Such apathy could be compounded by a lack or absence of senior leadership buy-in on using predictive LA, seriously impeding its effectiveness.[99] It is equally important that senior leaders are knowledgeable about predictive LA. Securing the endorsement of senior figures requires communication of the benefits of predictive LA. For instance, highlighting the literature that demonstrates how adopting predictive LA models has improved student retention and academic performance, as well as cost benefits, noting that student attrition has financial implications and costs in terms of time and institutional reputation.[100]

Serious thought must be given to the presentation of data generated by predictive LA models, as well as who has visibility. The design of a tool to present predictive LA model output should consider institutional levels of data literacy and be optimised to ensure easy understanding. Additionally, data presentation needs to be tailored to specific audiences. Dialogue and collaboration with stakeholders to ascertain what data points are required for their area of work is crucial. Stakeholders may include students as well as staff. As such, conversations with student representatives and personal tutors would be instructive in designing any data presentation tool. For example, Northumbria University developed an 'Integrated service dashboard' founded on the completion of a rapid evidence assessment and consultation with the Counselling and Mental Health Team. Susnjak, Ramaswami and Mathrani proposed a learner-facing analytics dashboard with descriptive, predictive, *and* prescriptive capabilities (Figure 4).

---

96 Huijser, West and Heath, "The Potential of Learning Analytics to Systematically Address Diverse Learning Needs and Improve Student Retention in Australian Higher Education."

97 D. West, H. Huijser, A Lizzio, D. Toohey, C. Miles, B. Searle, J. Bronnimann, *Learning Analytics: Assisting Universities with Student Retention, Institutional Analytics Case Studies* (2015); C. Colvin, T. Rogers, A. Wade, S. Dawson, D. Gasevic, S. Buckingham Shum, K. Nelson, S. Alexander, L. Lockyer, G. Kennedy, L. Corrin and J. Fisher, *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement* (Canberra, ACT, Australia: Australian Government – Office for Learning and Teaching, 2015).

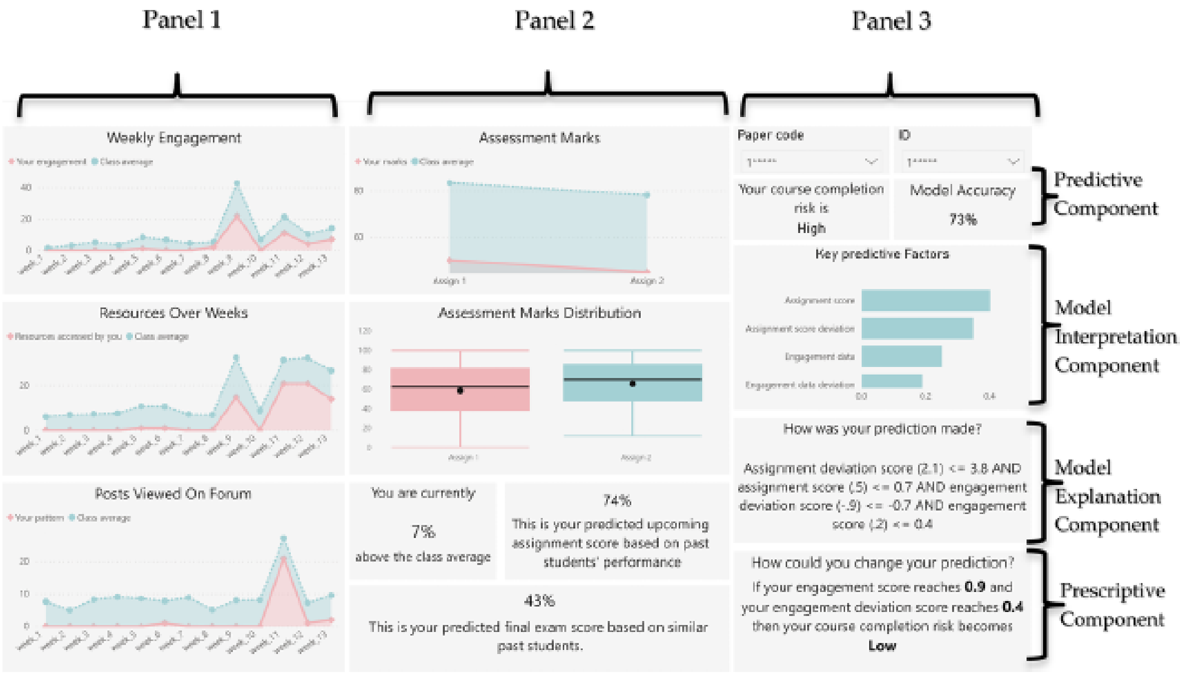
98 Huijser, West and Heath, "The Potential of Learning Analytics to Systematically Address Diverse Learning Needs and Improve Student Retention in Australian Higher Education."

99 Arroway *et al*., *Learning Analytics in Higher Education*.

100 Cochran *et al*., "The Role of Student Characteristics in Predicting Retention in Online Courses"; Arroway *et al*., *Learning Analytics in Higher Education*.

Features of this model include model interpretability, allowing students to understand the factors influencing predictions, and counterfactuals, which demonstrate specific changes students can make to improve outcomes.[101] Prescriptive analytics is an emerging area, and any output must be closely scrutinised. It is worth reiterating here that predictive/prescriptive LA tools are assistants; human input and expertise must be at the heart of decisions concerning signposting and interventions.

Figure 4. Learning analytics dashboard designed for students



5.0 Addressing Disengagement and Discontinuation

While student retention might be the primary driver for the development of predictive LA technologies, its application need not be limited to the early identification of at-risk students. HE providers should use the output of predictive LA models as a prompt to action. Indeed, numerous models have used the results from their models to guide their subsequent practices. For example, Herodotou *et al*. discussed a randomised control trial (RCT) undertaken involving 630 undergraduate students divided into control and intervention groups. The intervention group received motivational support via text, phone, and email. The intervention group showed statistically significant better retention outcomes compared to the control group.[102] The approach adopted at Northumbria University has been to send tailored "nudge" emails for students, with students being given the option at enrolment each year to consent to receiving

[101] Susnjak, Ramaswami and Mathrani, "Learning analytics dashboard."
[102] Herodotou *et al*., "How Can Predictive Learning Analytics and Motivational Interventions Increase Student Retention and Enhance Administrative Support in Distance Learning?"

these emails. While acknowledging that the COVID-19 pandemic and lockdowns likely contributed to the decrease in the number of self-referrals to support services, since the nudging of students based solely on their WHO-5 score began in 2021, rates of self-referral have shown an 11% increase on 2019 and 27% increase on 2020. Forecasts from the predictive model were used to judge an appropriate level of support for each student (Figure 5). Multiple factors influenced the effectiveness of a "nudge," including the communication channel and the message's clarity and brevity. However, the generalisability of the effectiveness of "nudge" emails *per se* is a subject in need of further research, as is their content.[103]

Figure 5. Judging the "nudge" level[104]

Risk level	**Nature of support**
Everyone	Generic signposting
High risk	One-to-one support
Medium risk	Guided self-help
Low risk	Wellbeing workshops

TASO reported on RCTs conducted by Nottingham Trent University and Sheffield Hallam University. The institutions tested two interventions: Intervention 1 – an email followed by a support phone call from a student adviser, and Intervention 2 – an email alert with details of available support. No measurable difference in engagement was found between students who received either intervention. Qualitative feedback from the project found that students welcomed the intervention. For some students, the phone call helped break down barriers between them and the institution, 'stimulating their re-engagement with learning'; others felt that the email was sufficient to motivate re-engagement.[105] Thus, there are several routes UoW could take in its efforts to reengage students. Importantly, each approach should be impact evaluated to ascertain the most effective remedial actions.

We are not advocating that a predictive LA model be left to its own devices, and space must remain for human supervision and decision-making. The report on Northumbria University's experience highlighted that despite employing predictive LA, deciding what to do with the data was labour-intensive – additional staff were appointed to meet the workload.[106] Thus, while part of the process may be automated and signposting becomes more sophisticated as the model

[103] Keane, *The Office for Students (OfS) mental health analytics project*.
[104] *Ibid*.
[105] TASO, *Using learning analytics to prompt student support interventions*, 3
[106] Keane, *The Office for Students (OfS) mental health analytics project*.

matures, in cases where human intervention is required, the judgement on the most appropriate course of action must reside with relevant staff and experts.[107]

Furthermore, insights from predictive models can guide educators and support staff in creating supportive interventions and curriculum development.[108] Student engagement is important for academic success, especially in the first year of study where discontinuation rates are highest.[109] There is an argument, therefore, that support should be front-loaded to assist students in the early stages of their university life.[110] Precisely what early these support interventions look like should reflect the needs of each student but may include personalised tutoring plans, instructional and timely feedback, tailored career guidance, support mechanisms such as drop-in sessions, online resources, and peer-assistant support.[111] There is also a role for targeted academic support, as well as financial assistance such as scholarships and bursaries.[112] Outreach has an important role in this regard. For example, one study found that fostering economic competencies at the upper secondary level was crucial for improving student retention in economics HE.[113] Concerning the curriculum, disengagement should prompt course and module leaders to consider how students can be best supported. Course content should be interactive and engaging, and a diverse range of pedagogical strategies (i.e., discussions, quizzes, practical activities, etc.) should be employed. In addition, support with academic skills, including writing, research and critical thinking, should be embedded into courses where possible.[114]

Conclusion

This paper has argued in favour of using AI to *assist* in detecting disengaged students and those at risk of discontinuing their studies. Indeed, the introduction stressed that predictive LA must only play an assistive role – it is not an alternative to human instinct or expertise. The REA assessed 56 papers concerning student attrition and/or student engagement, focusing on those employing ML techniques to identify engagement indicators. The review identified 48 indicators

---

107 Peck, *Student analytics*.

108 Huijser, West and Heath, "The Potential of Learning Analytics to Systematically Address Diverse Learning Needs and Improve Student Retention in Australian Higher Education"; Colvin *et al*., *Student retention and learning analytics*.

109 Paura and Arhipova, "Cause Analysis of students' dropout rate in higher education study program."

110 Beer *et al*., "Measuring engagement: An institution-wide implementation of learning analytics to increase retention"; and Aulck *et al*., "Predicting Student Dropout in Higher Education."

111 Foster and Francis. "A Systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes"; Wild, "Trajectories of subject-interests development and influence factors in higher education"; Dawson *et al*., "From prediction to impact: Evaluation of a learning analytics retention program"; Faridhan, Loch and Walker, "Improving retention in first-year mathematics using learning analytics"; Oliveira *et al*., "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education"; Sønderlund, Hughes and Smith, "The efficacy of learning analytics interventions in higher education."

112 Stylianou and Milidis, "The socioeconomic determinants of University dropouts."

113 Jüttler, "Predicting economics student retention in higher education."

114 Wild, "Trajectories of subject-interests development and influence factors in higher education"; Foster and Francis, "A Systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes"; Fan *et al*., "Supporting engagement and retention of online and blended-learning students"; Rotar, "A missing theoretical element of online higher education student attrition, retention, and progress: a systematic literature review."

spanning academic, attitudinal, demographic, behavioural and institutional arenas. Those tabulated constitute disengagement markers tested and used in other predictive LA models. They are unlikely to be exhaustive, and any procured model will need to factor in Westminster's distinct demographic and institutional context to pinpoint the most appropriate indicators and the balance of their importance. Evidence demonstrates that there is no single way to build a predictive LA model, though some statistical and ML techniques appear to offer greater precision and accuracy than others. In particular, multiple institutions reported high precision and accuracy scores for Random Forest techniques. This approach is of particular interest as it was recently used by Northumbria University to develop its predictive model, and we may wish to lean on this earlier and more advanced project for instruction and inspiration, keeping an eye on any impact evaluations. It is not only the technical aspects of model development that we may wish to seek instruction for. To be effective, predictive LA models rely on the supply of vast amounts of student data. As such, there are data protection and ethical questions to be navigated, as well as a large piece on data integration.

Bibliography

[**1**] Aina, Carmen, Eliana Baici, Giorgia Casalone and Francesco Pastore. "The determinants of university dropout: A review of the socio-economic literature." *Socio-Economic Planning Sciences* 79 (2022): 1-16.

[**2**] Al-Tameemi, Ghaith, James Xue, Suraj Ajit, Triantafyllos Kanakis, Israa Hadi. "Predictive Learning Analytics in Higher Education: Factors, Methods and Challenges." In: 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, June 2020): 1-9.

[**3**] Archambault, Isabelle, Michel Janosz, Elizabeth Oliver and Véronique Dupéré. "Student Engagement and School Dropout: Theories, Evidence, and Future Directions." In: *Handbook of Research on Student Engagement*, eds. A.L. Reschly and S.L. Christenson (Springer, 2002): 331-355.

[**4**] Arroway, Pam, Glenda Morgan, Molly O'Keefe and Ronald Yanosky. *Learning Analytics in Higher Education*. Research report. Louisville, CO: ECAR, March 2016.

[**5**] Asai, D.J. "Race Matters." *Cell* 181, no.4 (2020): 754-757.

[**6**] Aulck, Lovenoor, Nishant Velagapudi, Joshua Blumenstock and Jevin West. "Predicting Student Dropout in Higher Education." In: 2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications. https://arxiv.org/abs/1606.06364v4.

[**7**] Beer, Colin, Celeste Lawson, Gemma Mann and Damien Clark. "Measuring engagement: An institution-wide implementation of learning analytics to increase retention." CQUniversity. Conference contribution, 2017. https://hdl.handle.net/10018/1225043.

[**8**] Calvert, Carol Elaine. "Developing a model and applications for probabilities of student success: a case study of predictive analytics." *Open Learning: The Journal of Open, Distance and e-Learning* 29, no.2 (2014): 160-173.

[**9**] Cochran, Justin D., Stacy M. Campbell, Hope M. Baker and Elke M. Leeds. “The Role of Student Characteristics in Predicting Retention in Online Courses.” *Research in Higher Education* 55 (2014): 27-48.

[**10**] Colvin, C., T. Rogers, A. Wade, S. Dawson, D. Gasevic, S. Buckingham Shum, K. Nelson, S. Alexander, L. Lockyer, G. Kennedy, L. Corrin and J. Fisher. *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Canberra, ACT, Australia: Australian Government – Office for Learning and Teaching, 2015.

[**11**] Crane, Phoenix R., Katarina S. Swaringen, Matthew M. Rivas-Koehl, Anthony M. Foster, Tran H. Le, Dana A. Weiser, and Amelia E. Talley. “Come Out, Get Out: Relations Among Sexual Minority Identification, Microaggressions, and Retention in Higher Education.” *Journal of Interpersonal Violence* 37, no. 9-10 (2022): 8237-8248.

[**12**] Dawson, Shane, Jelena Jovanovic, Dragan Gašević, and Abelardo Pardo. "From prediction to impact: Evaluation of a learning analytics retention program." In *Proceedings of the seventh international learning analytics & knowledge conference*, pp. 474-478. 2017.

[**13**] Denaro, Kameryn, Kimberly Dennin, Michael Dennin and Brian Sate. “Identifying systemic inequity in higher education and opportunities for improvement.” *PLoS ONE* 17, no.4 (2022): 1-16.

[**14**] Fan, Si, Allison Trimble, David Kember, Tracey Muir, Tracy Douglas, Yanjun Wang, Jennifer Masters and Casey Mainsbridge. “Supporting engagement and retention of online and blended-learning students: A qualitative study from an Australian University.” *The Australian Education Researcher* 11 (2023): 1-19.

[**15**] Faridhan, Yasmin Erika, Birgit Loch and Lyndon Walker. “Improving retention in first-year mathematics using learning analytics.” In: Electric Dreams, 30th ASCILITE Conference, Macquarie University, Sydney, 1-4 December 2013.

[**16**] Foster, Carly and Peter Francis. “A Systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes.” *Assessment and Evaluation in Higher Education* 45, no.6 (2020): 822-841.

[**17**] Freitas, Sara de, David Gibson, Coert Du Plessis, Pat Halloran, Ed Williams, Matt Ambrose, Ian Dunwell and Sylvester Arnab. “Foundation of dynamic learning analytics: Using university student data to increase retention.” *British Journal of Educational Technology* 46, no.6 (2015): 1175-1188.

[**18**] Guzmán-Valenzuela, Carolina Gómez González, Andrés Rojas-Murphy Tagle and Alejandro Lorca-Vyhmeister. “Learning analytics in higher education: a preponderance of analytics but very little learning?” *International Journal of Educational Technology in Higher Education* 18, no.23 (2021): 1-19.

[**19**] Herodotou, Christothea, Galina Naydenova, Avi Boroowa, Alison Gilmour and Bart Rienties. “How Can Predictive Learning Analytics and Motivational Interventions Increase Student Retention and Enhance Administrative Support in Distance Learning?” *Journal of Learning Analytics* 7, no.2 (2020): 72-83.

[**20**] Higher Education Funding Council for Wales. *Use of data to support Student Engagement in Higher Education*. HEFCW, 2024.

[**21**] Huijser, Henk, Deborah West and David Heath. "The Potential of Learning Analytics to Systematically Address Diverse Learning Needs and Improve Student Retention in Australian Higher Education." *Advances in Scholarship of Teaching and Learning* 13, no.1 (2016): 1-19.

[**22**] Jia, Pengfei, and Tim Maloney. "Using predictive modelling to identify students at risk of poor university outcomes." *Higher Education* 70 (2015): 127-149.

[**23**] Jüttler, Michael. "Predicting economics student retention in higher education: The effects of students' economic competencies at the end of upper secondary school on their intention to leave their studies in economics." *PLOS ONE* 15, no.2 (2020): 1-27.

[**24**] Kahu, Ella R. "Framing student engagement in higher education." *Studies in Higher Education* 38, no.5 (2013): 758-773.

[**25**] Keane, Jim. *The Office for Students (OfS) mental health analytics project: an evaluation*. (Jisc, 2024).

[**26**] Lee, Lap-Kei, Simon K.S. Cheung and Lam-For Kwok. "Learning analytics: current trends and innovative practices." *Journal of Computers in Education* 7, no.1 (2020): 1-6.

[**27**] Li, Chunping, Nicole Herbert, Soonja Yeom and James Montgomery. "Retention Factors in STEM Education Identified Using Learning Analytics: A Systematic Review." *Education Sciences* 12, no.781 (2022): 1-18.

[**28**] Marôco, João, Hugo Assunção, Hiedi Harju-Luukkainen, Su-Wei Lin, Pou-Seong Sit, Kwok-cheung Cheung, Benvindo Malao, Ivana Stepanović Ilic, Thomas J. Smith and Juliana A.D.B. Campos. "Predictors of academic efficacy and dropout intention in university students: Can engagement suppress burnout?" *PLOS ONE* 15, no.10 (2020): 1-26.

[**29**] Matz, Sandra C., Christina S. Bukow, Heinrich Peters, Christine Deacons, Alice Dino and Clemens Stachl. "Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics." *Nature: Scientific Report* 13, 5705 (2023): 1-16.

[**30**] Nurmalitasari, Zalizah Awang Long and Mohammad Faizuddin Mohd Noor. "Factors Influencing Dropout Students in Higher Education." *Education Research International* (2023): 1-13.

[**31**] Oliveira, Catarina Félix de, Sónia Rolland Sobral, Maria João Ferreira and Fernando Moreira. "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education: A Systematic Review." *Big Data and Cognitive Computing* 5, no.64 (2021): 1-33.

[**32**] Ortiz-Martínez, Gabriela, Patricia Vázquez-Villegas, María Ileana Ruiz-Cantisani, Mónica Delgado-Fabián, Danna A. Conejo-Márquez and Jorge Membrillo-Hernández. "Analysis of the retention of women in higher education STEM programs." *Humanities and Social Sciences Communications* 10, no.101 (2023): 1-14.

[**33**] Parkes, Sarah, Adam Benkwitz, Helen Bardy, Kerry Myler and John Peters. "Being more human: rooting learning analytics through *re*sistance and *re*connection with the values of higher education." *Higher Education Research & Development* 39, no.1 (2020): 113-126.

[**34**] Paterson, Kevin and Adam Guerrero. "Predictive Analytics in Education: Considerations in Predicting versus Explaining College Student Retention." *Research in Higher Education Journal* 44 (2023): 1-12.

[**35**] Paura, Liga, and Irina Arhipova. “Cause Analysis of students’ dropout rate in higher education study program.” *Procedia – Social and Behavioural Sciences* 109 (2014): 1282-1286.

[**36**] Peck, Edward. *Student analytics: A core specification for engagement and wellbeing analytics*. Jisc, 2023.

[**37**] Pedler, Megan Louise, Royce Willis and Johanna Elizabeth Nieuwoudt. “A sense belonging at university: student retention, motivation and enjoyment.” *Journal of Further and Higher Education* 46, no.3 (2022): 397-408.

[**38**] Raj, Nisha S. and V.G. Renumol. “Early prediction of student engagement in virtual learning environments using machine learning techniques.” *E-Learning and Digital Media* 19, no.6 (2022): 537-554.

[**39**] Rawlinson, Scott. “Academic and Practical Information Seeking Behaviours and Needs of International Students at Pre-arrival and Arrived (First Year) Stages.” Westminster – University of Westminster, 2023.

[**40**] Rawlinson, Scott and Jo Alexander. “Perspectives on Student (Dis)Engagement and Continuation (Academic and Support Colleagues).” Westminster: University of Westminster – Institutional Research, 2023.

[**41**] Rawlinson, Scott. “‘Little Islands’: challenges and opportunities for student carers in higher education.” *International Journal of Inclusive Education* (2024): 1-16.

[**42**] Realinho, Valentim, Jorge Machado, Luis Baptista and Monica V. Martins. “Predicting Student Dropout and Academic Success.” *Data* 7, no.146 (2022): 1-17.

[**43**] Rodríguez-Muñiz, Luis J., Ana B. Bernardo, María Esteban and Irene Díaz. “Dropout and transfer paths: What are the risky profiles when analyzing university persistence with machine learning techniques?” *PLOS ONE* 14, NO.6 (2019): 1-20.

[**44**] Rotar, Olga. “A missing theoretical element of online higher education student attrition, retention, and progress: a systematic literature review.” *SN Social Sciences* 2, no.12 (2022): 277-278.

[**45**] Sclater, Niall and Joel Mullan. *Learning analytics and student success – assessing the evidence*. Jisc, 2017.

[**46**] Seidel, Ewa and Salah Kutieleh. “Using predictive analytics to target and improve first year student attrition.” *Australian Journal of Education* 61, no.2 (2017): 200-218.

[**47**] Sghir, Nabila, Amina Adadi and Mohammed Lahmer. “Recent advances in Predictive Learning Analytics: A decade systematic review.” *Education and Information Technologies* 28 (2023): 8299-8333.

[**48**] Shafiq, Dalia Abdulkareem, Mohsen Marjani, Riyaz Ahamed Ariyaluran Habeeb, and David Asirvatham. "Student retention using educational data mining and predictive analytics: a systematic literature review." *IEEE Access* 10 (2022): 72480-72503.

[**49**] Shaikh, Umair Uddin and Zaheeruddin Asif. “Persistence and Dropout in Higher Online Education: Review and Categorisation of Factors.” *Frontiers in Psychology* (2022).

[**50**] Sønderlund, Anders Larrabee, Emily Hughes and Joanne Smith. "The efficacy of learning analytics interventions in higher education: A systematic review." *British Journal of Educational Technology* 50, no.5 (2019): 2594-2618. doi:10.1111/bjet.12720.

[**51**] Stylianou, Tasos and Alexandros Milidis. "The socioeconomic determinants of University dropouts: The case of Greece." *Journal of Infrastructure, Policy and Development* 8, no.6 (2024): 1-21.

[**52**] Susnjak, Teo, Gomathy Suganya Ramaswami and Anuradha Mathrani. "Learning analytics dashboard: a tool for providing actionable insights to learner." *International Journal of Educational Technology in Higher Education* 19, no.12 (2022): 1-23.

[**53**] TASO, *Using learning analytics to prompt student support interventions: Findings from two randomised controlled trials* (TASO: February 2024).

[**54**] West, D., H. Huijser, A Lizzio, D. Toohey, C. Miles, B. Searle, J. Bronnimann. *Learning Analytics: Assisting Universities with Student Retention, Institutional Analytics Case Studies* (2015).

[**55**] Wild, Steffen. "Trajectories of subject-interests development and influence factors in higher education." *Current Psychology* 42 (2022): 12879-12895.

[**56**] Wolff, Annika, Zdenek Zdrahal, Andriy Nikolov, and Michal Pantucek. "Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment." In: *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, Association for Computing Machinery, New York, NY, USA, pp. 145–149, 2013.

Appendix 1. The World Health Organisation-Five Well-Being Index (WHO-5)

The World Health Organisation-Five Well-Being Index (WHO-5)

Please indicate for each of the five statements which is closest to how you have been feeling over the last two weeks. Notice that higher numbers mean better well-being.

Example. If you have felt cheerful and in good spirits more than half of the time during the last two weeks, select number three.

		All of the time	Most of the time	More than half of the time	Less than half of the time	Some of the time	At no time
1	I have felt cheerful and in good spirits	5	4	3	2	1	0
2	I have felt calm and relaxed	5	4	3	2	1	0
3	I have felt active and vigorous	5	4	3	2	1	0
4	I woke up feeling fresh and rested	5	4	3	2	1	0
5	My daily life has been filled with things that interest me	5	4	3	2	1	0