26TH INTERNATIONAL CONFERENCE ON ENGINEERING AND PRODUCT DESIGN EDUCATION
5-6 SEPTEMBER 2024, ASTON UNIVERSITY, BIRMINGHAM, UNITED KINGDOM

DEVELOPING AN OPEN-SOURCE LEARNING ANALYTICS TOOL FOR PROVIDING INSIGHTS TO SUPPORT STUDENTS AND IMPROVE TEACHING PRACTICE

Derek COVILL, James TOOZE, Pablo PRIETO, Gareth OWEN LLOYD and Cate GRUNDY
University of Brighton, United Kingdom

ABSTRACT
Learning analytics and data-driven approaches have gained real traction in higher education institutions, enabling insights into student attendance, engagement, and performance trends. This paper introduces the development of an open-source course-level data analytics tool, 'the student record,' designed to complement the conventional institutional approach with a responsive, context-specific, and user-centred method for gathering, analysing, and presenting student attendance data. In developing the tool, our team experimented with Microsoft Excel's natural language interface and artificial intelligence capabilities and currently the tool uses predefined python scripts in a Jupyter Notebook to generate visual dashboards, flags, and recommendations based on statistical process control (SPC) principles and machine learning models. This approach aims to facilitate a dynamic student engagement dialogue between staff and students, fostering a rolling academic health check to support students and provide contextual insights for module teams and course leaders. The tool's ethical considerations prioritise transparency, data privacy, and student accessibility to their own data, aligning with personal academic tutoring systems. The paper outlines the tool's development, capabilities, and a critical discussion of key aspects, including AI integration, data security, and ethical implications.

Keywords: student attendance and engagement, machine learning, ethics, information system, contextual insights.

1 INTRODUCTION AND BACKGROUND
Data analytics has been employed extensively in the higher education sector in recent years. It has enabled powerful insights into trends such as the growing gap between male and female applicants in England [1], attainment gaps between ethnicity groups [2, 3], the rise in the number of students reporting mental health issues [4], and predictions of the rise and fall in the populations of 18-year-olds [5] as shown in Figure 1.

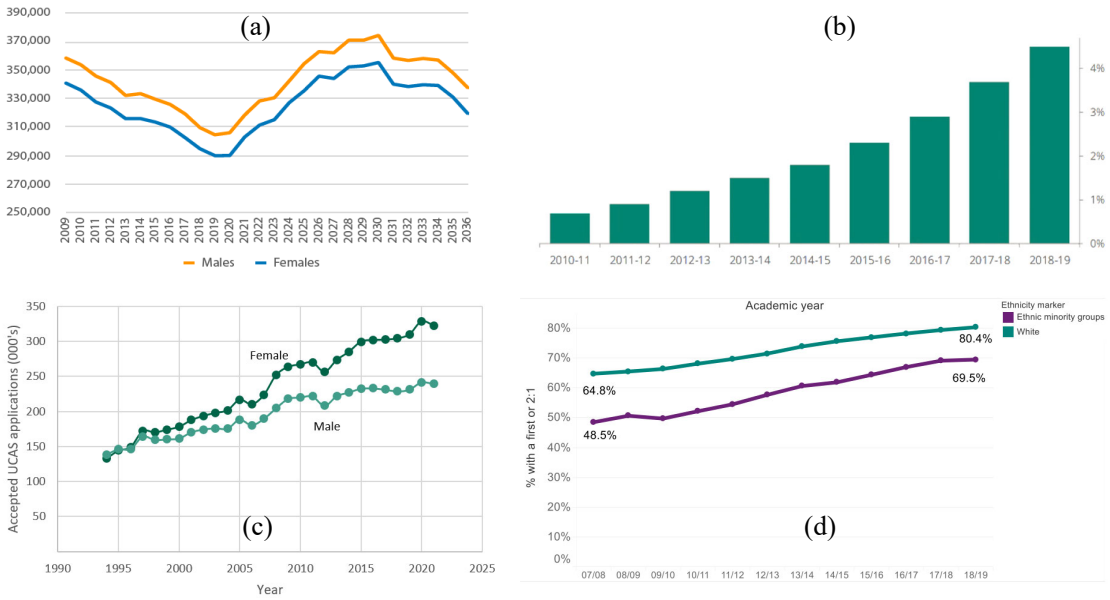


Figure 1. a) English and Welsh 18-year-old population predictions to 2036 [1]

b) Proportion of students reporting a mental health condition in English Universities [2], c) Male and female accepted applicants in UK Universities [3], d) Percentage first degree qualifiers achieving top degrees by ethnicity [4]

A more student-centred approach to data analytics is learning analytics, which can be a valuable tool for gaining insights into general trends to help course teams develop, but also a means to provide more targeted support to students at course, module, and individual levels. At our institution, we have a generic web-based register tool for reporting data into a central visualisation tool (QlikView), but these are not designed for reporting and have real limitations for extracting useful insights. Our web-register tool also only provides data for the most recent two-week window, and so it fails to give a complete picture over a semester or academic year. In addition, there is a system (StudentView) which gives completed module results, however this only provides a single final result for modules. Overall, these systems do not work together in an integrated way and leaving little room for transparency or student ownership of engagement or attendance data. While results and feedback are contained within the virtual learning environment alongside teaching content, personal tutors and staff who do not teach on those modules do not have access to student outcomes, and so there is no integrated solution that can inform personal tutor conversations with students. Furthermore, it seems design and other studio-based courses do not seem well serviced by typical attendance monitoring systems. As a result, we were keen to develop a tool that would service the needs for our studio-based product design and design engineering courses in order to be better informed in our support for students.

In terms of monitoring attendance in a university setting, Newman-Ford et al. [6] have provided compelling evidence of a significant relationship between attendance and attainment in a large-scale study involving 22 undergraduate modules (Figure 2). They also showed that self-developed tools can be a powerful way to gain high-level and detailed insights that can cater to specific institutional or departmental needs. While there is a current trend for more streamlined, technological means to gather data (e.g., face recognition, video, fingerprint, RFID), there seems to be less focus on using technology to generate useful insights to directly support students or inform and improve teaching practices.

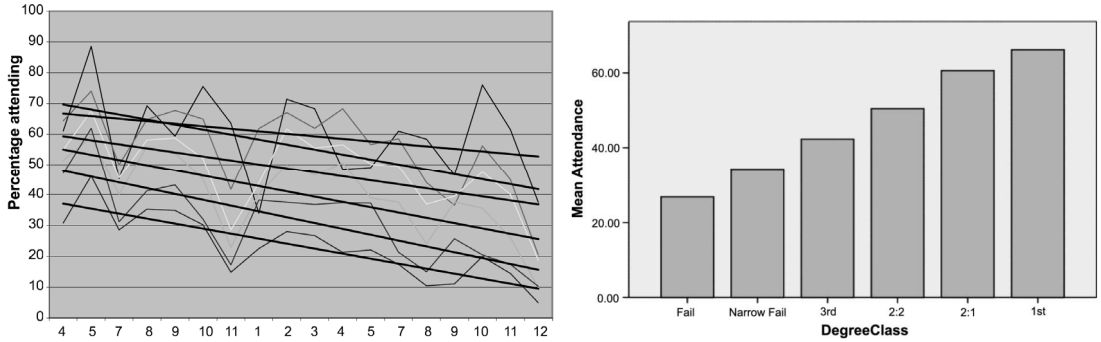


Figure 2. Attendance across two terms by assessment result (left) and attendance by performance (right) [6]

Statistical process control (SPC) is a long-standing total quality management approach that provides a framework for identifying patterns and interpreting data, enabling statistical inference and assessment, and generating insights and this has also been applied successfully to an education setting to improve outcomes [7]. Total quality management (TQM) processes emphasise continuous improvement by minimising failures, and it is common practice for institutions to present visual information through staff or student-facing dashboards to report on attendance, engagement and attainment. By extension, one study has even experimented with using a 'risk of failure' as part of visualisation dashboard integrated into a student counselling system to inform academic staff when advising students as shown in Figure 3 [8].

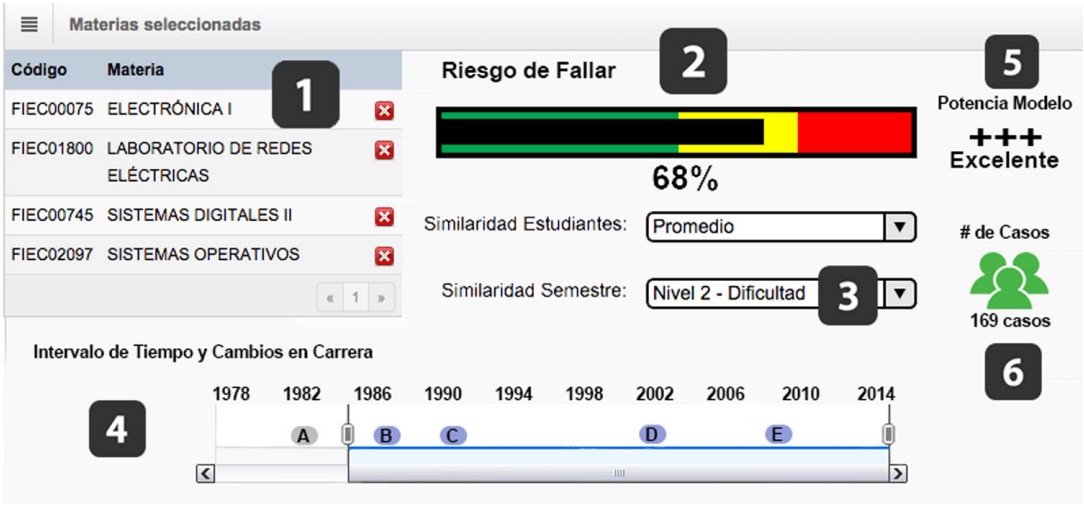


Figure 3. Example of visualisations integrated into a student counselling system: 1) Course selector, 2) Predicted academic risk value visualisation, 3) Model selector, 4) Time period selector and consistency visualisation, 5) Model predictive power visualisation and 6) Case completeness visualisation [7]

2 DEVELOPMENTS OF A COURSE LEVEL LEARNING ANALYTICS SYSTEM

2.1 System requirements

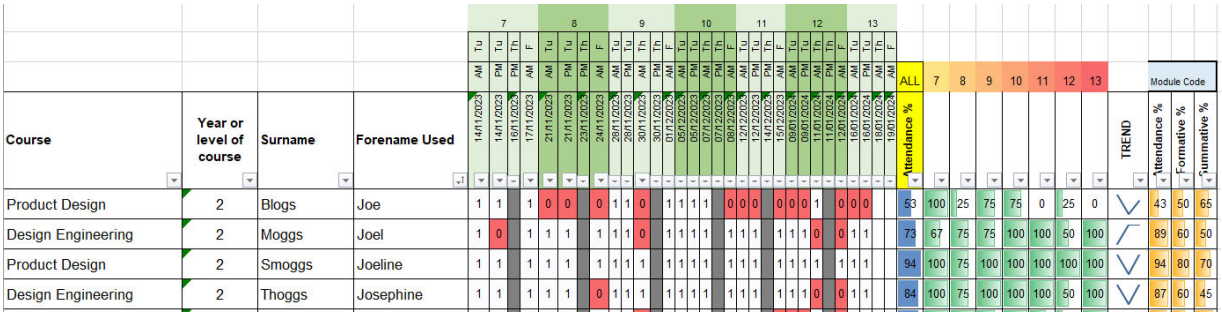
Based on our identified needs and objectives, a set of user requirements and corresponding specifications were outlined for the development of the course-level learning analytics system as shown in Table 1.

Table 1. System requirements and specifications

Requirements	Specification
Provide a visual dashboard to display student attendance, engagement, and assessment data.	The system should generate configurable dashboards with graphical representations of student data, including attendance percentages, formative and summative assessment results, and feedback.
Foster transparency and student ownership of their data.	The system should allow students to access and view their own data, aligning with personal academic tutoring systems.
Generate insights for students, staff, module leaders, course leaders, and the institution.	The system should provide targeted recommendations and actionable insights tailored to different stakeholder groups. User Requirement: Facilitate targeted action to support students.
Facilitate targeted action to support students.	The system should identify students at risk or in need of support based on their attendance, engagement, and performance data, enabling timely interventions and personalised support.
Ensure data security and ethical considerations.	The system should prioritise data privacy and adhere to ethical guidelines, including transparency in data handling and secure storage.

2.2 Data collection for student attendance, engagement and assessment

Data collection for student attendance, engagement, and assessment was aligned with the existing workflow of the institution's Web Register tool to simplify the process for staff when inputting data. A password-protected Excel file including tabs for each module and for summary dashboards was located on the course team's secure SharePoint site. In this file included information with the following separate columns 1) course, 2) year, 3) student number, 4) Surname, 5) Firstname, 6) formative assessment mark, 7) formative feedback comments, 8) summative assessment mark/s, 9) summative assessment mark/s, 10+) attendance for each session (separate columns for each) 11) module attendance, 12) notes relating to module as shown in Figure 4 which shows an excel template with filters and conditional formatting.



Course	Year or level of course	Surname	Forename Used	Attendance %	TREND	Attendance %	Formative %	Summative %
Product Design	2	Blogs	Joe	53		43	50	65
Design Engineering	2	Moggs	Joel	73		89	60	50
Product Design	2	Smoggs	Joeline	94		94	80	70
Design Engineering	2	Thoggs	Josephine	84		87	60	45

Figure 4. An example student record dataset for a module (fictitious)

2.3 A student record: Central to the ongoing 'personal tutoring' dialogue

The 'student record' could then serve as the main record for conversations between students and their personal tutors, module tutors, course leaders, and student support and guidance tutors. The key aspect is to present information visually and engage students in analysing their data. For each module, the student record includes attendance data (cumulative percentage), formative assessment results and feedback, and summative assessment results and feedback. Attendance data is able to be presented in multiple formats, including self-comparison (e.g. comparing attendance in multiple modules, or comparison between morning and afternoon sessions), comparison to others on the module (anonymous collective data), and comparison to the entire course cohort. Metrics and trend analyses are conducted, such as fluctuations, upward/downward trends, percentage above a threshold in a rolling period (e.g. every two weeks) in line with SPC run charts which can help to flag statistically significant scenarios that deviate from the ordinary flow of events for modules or individual students. The Filter tool in Excel enables the presentation of data for individual students during personal tutorial meetings, and also the display of groups of students with low attendance or attainment in formative or summative assessments for module or course level analysis. Conditional formatting is employed to colour-code attendance and

attainment data, aiding visualisation and Macros were used to improve the usability of the system by including programmed buttons for searching, sorting and filtering data for students, modules and courses.

A more sophisticated system was then developed as a Jupyter notebook which processes a series of Python scripts to compare attendance patterns (e.g. across different days, or between morning and afternoon sessions) are shown as box plots and are analysed to highlight potential issues with sleep patterns, commuting difficulties, employment, or caring responsibilities. In response to some requests from colleagues in other departments, an open-source repository of the student record system was made available on Github at https://github.com/derekcovill/student-record/).

Figure 5 (left) below shows an example plot showing attendance for a module for an individual student. The analysis of module and course-level data allows for a comparison of attendance and attainment data across modules and courses presented as quartile data (upper, median, lower as shown in Figure 4 - right). This goes beyond the formal data reported centrally at exam boards, offering a more comprehensive view of student performance and engagement for module and course teams. The work is currently being extended to exploit the natural language AI insights tool (*Analyse Data*) and GPT plug-ins for Excel. For students, this tool can potentially be used to generate insights into their individual behaviour, while module leaders can examine students' collective behaviour, and course leaders can compare data across modules. The AI tools within Excel should also allow us to extract trends and common traits in formative and summative feedback, packaging these for staff to provide guidance and further support to students. This can be done within a module, across modules, across years for modules, and for individual students to synthesise comments across modules and identify key areas for improvement.

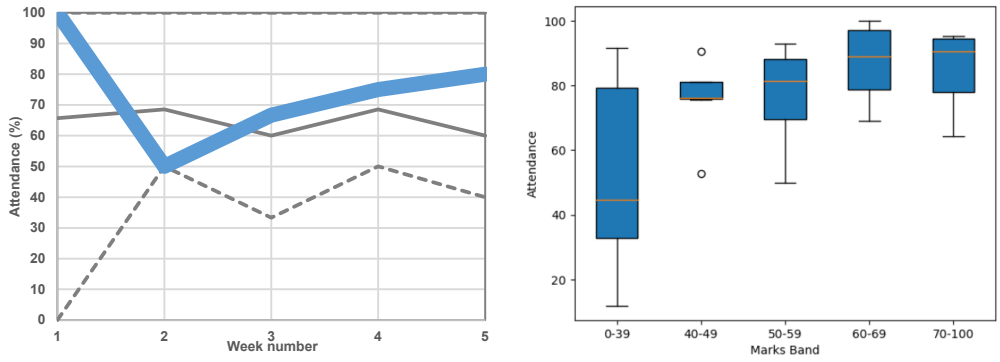


Figure 5. Example plot showing attendance for a module for a student (thick blue line), with dashed grey lines showing the upper and lower quartiles and the solid grey line the median student percentage (left) and example box plot showing percentage results for various grade bands (right)

2.4 Course and module dashboard: Analysis of collective student behaviour and course oversight

The course and module dashboard provides a comparison of attendance and attainment data across modules and courses presented as quartile data (upper, median, lower as shown in Figure 4 - right). This goes beyond the formal data reporting that occurs centrally at exam boards, offering a more comprehensive view of student performance and engagement.

Currently some testing is being done to further exploit the natural language AI insights tool (*Analyse Data*) and using machine learning tools in our Jupyter notebook using regression and classification models for predicting outcomes and identifying at-risk groups based on their engagement and attainment profiles. For students, the tool can be used to generate insights into their individual behaviour, while module leaders can examine students' collective behaviour, and course leaders can compare data across modules. The more sophisticated tools seem capable of extracting trends and common traits in feedback, packaging these for staff to provide guidance and further support to students. This can be done within a module, across modules, across years for modules, and for individual students to synthesise comments across modules and identify key areas for improvement.

3 FINAL REMARKS

Our course-level data analytics tool, 'the student record,' has started to provide some data-driven insights using various technologies to support student success in design-related courses. By integrating attendance, engagement, and assessment data in a way that is accessible to the course team, the system provides a useful view of student performance and enables targeted interventions and support. The system's key strengths lie in its visual dashboards, trend analyses, and the integration of data across the course into a robust system that is easy to use for the course team. These features facilitate a dynamic 'student engagement' dialogue, allowing staff to identify potential issues and provide tailored guidance to students. Ethical considerations, such as transparency, data privacy, and student accessibility to their own data, have been prioritised in a relatively simple system design, aligning with personal academic tutoring systems and fostering a reflective learning process for staff and students. While the system has been developed to some extent through experimentation and exploration in the past 12 months to meet our own needs, further exploration of other machine learning tools and the scalability across different departments or even institutions could enhance its capabilities and impact.

REFERENCES

[1] Bolton P. *Higher education student numbers*. House of commons library. 2024.

[2] Universities UK (UUK). *Black, Asian and Minority Ethnic Student Attainment at UK Universities*, 2019.

[3] Joice P. and Tetlow G. *Ethnicity STEM data for students and academic staff in higher education 2007/08 to 2018/19*. A royal society report, 2020.

[4] Hubble S. and Bolton P. *Mental health and well-being among students at UK universities*. House of Commons Library, 2022

[5] Hewitt R. Demand for higher education to 2035. Higher Education Policy Institute (HEPI), report 134, 2020.

[6] Newman-Ford L., Fitzgibbon K., Lloyd S. and Thomas S. A large-scale investigation into the relationship between attendance and attainment: a study using an innovative, electronic attendance monitoring system. *Studies in Higher Education*, 2008, 33(6), 699-717.

[7] Marsh J. Quality in education on the move: Using Statistical Process Control to Improve Attendance. *Quality Digest Magazine*. 1995, 15(9), pp.26-32.

[8] Ochoa X. 2015, March. *Visualising Uncertainty in the Prediction of Academic Risk*. In VISLA@ LAK (pp. 4-10).