



An Extended Learning Analytics Framework Integrating Machine Learning and Pedagogical Approaches for Student Performance Prediction and Intervention

Khalid Alalawi^{1,2} · Rukshan Athauda¹  · Raymond Chiong^{1,3}

Accepted: 25 August 2024
© The Author(s) 2024

Abstract

The use of educational data mining and machine learning to analyse large data sets collected by educational institutions has the potential to discover valuable insights for decision-making. One such area that has gained attention is to predict student performance by analysing large educational data sets. In the relevant literature, many studies have focused on developing prediction models on student performance but comparatively less work exists on actions taken based on predicted at-risk students and evaluating their impact. Learning Analytics Intervention (LAI) studies have emerged as an approach that aims to address this gap. In LAI studies, student risk levels are predicted and disseminated to relevant stakeholders (academics, administrators and students) using learning analytics (LA) tools for targeted interventions. The interventions themselves are mainly left under the discretion of the academics and/or administrators, who are aware of the learning context and have the authority to make decisions, with LA tools facilitating this process. LAI studies have shown success in improving outcomes (e.g. improve pass rates, retention, grades), but their uptake has been slow. The main impediment to piloting LAIs by academics has been the lack of access to LAI infrastructure, which requires institutional investments to develop predictive models collecting data from diverse IT systems. Another challenge in LAIs is the development of effective interventions. This paper presents an extended LAI framework, termed Student Performance Prediction and Action (SPPA), which provides access to LAI infrastructure for academics to pilot LAIs in their courses without the need for institution-wide efforts. SPPA and its features are seamlessly accessed via a web browser and academics can develop course-specific predictive models based on historical course assessment data. Furthermore, SPPA integrates sound pedagogical approaches and provides relevant information (such as students' knowledge gaps, personalised study plans) to assist academics in providing effective interventions. SPPA was evaluated by a number of academics piloting LAIs in their courses. Quantitative and qualitative data was collected and analysed. Academics were able to provide effective interventions using SPPA and also had a positive outlook on using SPPA and its features. SPPA is also provided as

Extended author information available on the last page of the article

an open-source project for further development and can be a catalyst for widescale uptake in LAIs. Furthermore, a model for continuous improvement in LAIs is outlined along with a number of areas for future research and development.

Keywords Learning analytics interventions · Student performance prediction · Interventions for at-risk students

Introduction

Educational institutions these days use a variety of information systems, such as Learning Management Systems (LMS), Student Information Systems (SIS) and others, which collect vast quantities of data. These data sets, if analysed, can provide a trove of information to assist in decision-making to improve educational outcomes. Technological advances such as Artificial Intelligence (AI), Machine Learning (ML), Data Mining (DM), Data Analytics (DA) and others have enabled the analysis of such data sets to provide useful and actionable insights. Educational institutions are dedicating significant resources to analyse such data sets and develop analytical tools to make informed decisions (Baneres et al., 2019; Schwendimann et al., 2016). Educational Data Mining (EDM) and Learning Analytics (LA) are some of the fields that aim to leverage data to improve educational outcomes and offer promising avenues for institutions to support students effectively (Manjarres et al., 2018; Mohamad & Tasir, 2013).

An area that has gained significant research interest is the use of the educational data sets to predict student performance. There has been considerable work on predicting student performance using EDM techniques analysing these large datasets (Begum & Padmannavar, 2022; Enaro & Chakraborty, 2020; Gajwani & Chakraborty, 2021; Garg et al., 2021; Mubarak et al., 2022; Ramaswami et al., 2022; Song et al., 2023). These approaches have demonstrated success in predicting students' academic performance accurately as well as identifying at-risk students. Identifying at-risk students allows academics and institutions' support services to provide personalised interventions resulting in better educational and student outcomes. In the literature, we observe that comparatively few studies have investigated taking interventions based on predicted at-risk students and evaluating their impact. Learning Analytics Intervention (LAI) studies aim to address this gap (Larrabee Søndelund et al., 2019; B. T.-m. Wong & Li, 2020; Foster & Francis, 2020). In LAIs, predictive models are developed using data gathered from different educational information systems (SIS, LMS and others) that identifies student academic risk levels early on and disseminates this information to stakeholders (academics, administrators and students) to take appropriate actions while also providing LA tools to facilitate targeted interventions. Many LAI studies have shown positive results due to targeted interventions such as improved pass rates, improved grades and reduced dropout rates (Arnold & Pistilli, 2012; Baneres et al., 2019, 2023; Borrella et al., 2022; Burgos et al., 2018; Figueroa-Cañas & Sancho-Vinuesa, 2021; Jayaprakash et al., 2014; Milliron et al., 2014).

Although LAI studies have shown promise of improved outcomes since 2012 (Arnold & Pistilli, 2012), we are yet to see widespread uptake and adoption of LAIs in higher education. This paper contributes to the field of LAIs in a number of ways. Firstly, the paper analyses the literature and identifies a number of challenges and inhibitions for the uptake of LAIs. Next, the paper presents a framework that builds on the LAI approach, termed Student Performance Prediction and Action (SPPA), aiming to address the gaps in the uptake of LAIs. SPPA is evaluated by academics to provide LAIs in a number of courses in a tertiary education environment. Analysis of the results demonstrated SPPA's ability to provide impactful interventions. Also, academics had a positive outlook using SPPA. SPPA addressed a major obstacle that inhibits academics from piloting LAIs in their courses, which is the lack of access to LAI infrastructure, as it requires substantial investment by the educational institution. SPPA provided a self-service model for academics to pilot LAI in their courses, allowing seamless access to its features via a web browser without the need for extensive efforts in developing institution-centric LAI infrastructure. Finally, SPPA's approach was also extended to propose a continuous improvement model in LAIs.

The rest of this paper is organised as follows: In the next section, we discuss LAI studies in the LAI literature. After discussing the related work, we identify gaps and challenges in LAIs and establish research questions for this study. We then derive sub research questions from the main research questions and outline the study's design to address these research questions. Having done so, we present the SPPA framework. SPPA was evaluated using several courses. Some discussion on results and findings, limitations of the study, and potential areas for future work is subsequently presented. Finally, we conclude the paper.

Related Work

Many studies in the literature have focused on using EDM techniques, particularly classification algorithms, to predict students' performance, identify students at risk of failure or dropping out (Alalawi et al., 2023; Begum & Padmannavar, 2022; Enaro & Chakraborty, 2020; Gajwani & Chakraborty, 2021; Garg et al., 2021; Mubarak et al., 2022; Ramaswami et al., 2022; Song et al., 2023). Typically, data sets and features utilised for prediction include student performance data, student engagement data, and student demographic data. The ability to predict students' academic performance and identify at-risk students early is critical to implement effective interventions focusing on at-risk students to improve educational outcomes (Kovacic, 2010; Sclater et al., 2016; Wong, 2017). However, despite extensive research in this area, limited previous work has looked into the implementation of interventions based on prediction results and evaluating their impact (Larrabee Sønderlund et al., 2019).

Purdue University's Course Signals (CS) project (Arnold & Pistilli, 2012) is an early higher education initiative that employs predictive models to detect students who are at risk based on four sources of data: current course grades, LMS engagement, prior academic performance, and demographic information. The system

divides students into three groups: red, yellow, and green. A red signal indicates that a student is likely to fail. The yellow signal indicates that a student might fail, while the green signal indicates that a student is likely to pass their course. Based on students' signal results, academics can create an intervention strategy that possibly consists of posting a student's signal into the student's home page in the LMS, sending personalised emails or text messages, referral to an academic advisor or academic resource centres, or scheduling a face-to-face meeting. The CS project was implemented at Purdue University in 2007, 2008 and 2009 cohorts to analyse its impact on students' performance and retention rates (Arnold & Pistilli, 2012). The project has demonstrated a significant improvement in success and retention rates across many cohorts at Purdue.

The Open Academic Analytics Initiative (OAAI) project by Jayaprakash et al. (2014) takes a similar approach to CS. They evaluated different prediction algorithms and portability of them across institutions. This OAAI project considers four data sources: student demographics, course grades and related data, student interaction data with the LMS, and students' progress on their final grades so far. The goal of predictive analysis is to classify a student in good standing or academically at risk. Predictive models considered four well-known ML algorithms and evaluated at 25%, 50% and 75% of the semester. Logistic regression was chosen as the model of choice as it outperformed other algorithms. Next, this predictive model was ported across institutions and evaluated with interventions. The study evaluated two intervention strategies termed "Awareness Messaging Intervention" and "Online Academic Support Environment Intervention". The "Awareness Messaging Intervention" group received a message indicating that they were at risk of not completing the course successfully along with instructors encouraged to recommend actions (such as visit during office hours, set up an appointment with the tutor, access web-based resources, etc.). The "Online Academic Support Environment Intervention" group received a similar message to the "Awareness Intervention" group, except that instead of specific recommendations, the students were encouraged to join the institution-wide online support services – such as a range of mentoring from peers and professional support staff and also access to Open Educational Resources (OER) instructional materials (e.g., Khan Academy videos, Flat World Knowledge textbooks, etc.). Evaluation of control vs intervention groups showed a statistically significant difference in final grades in the control group vs intervention groups with a 6% improvement in final grades in the two treatment groups vs a non-intervention control group. There was no statistical significance difference between the two intervention groups. For at-risk students designated as lower socio-economic status (SES), a 7% increase in final grades was observed. In terms of withdrawal rates, 25.6% of students receiving the intervention dropped out compared to only 14.1% of the control group. The researchers speculated that this might have been due to students opting to discontinue their studies rather than failing at the end of semester.

Milliron et al. (2014) described three case studies where Civitas Learning's Illume platform is used to predict and provide interventions for at-risk students in three different institutions. In the case studies, data from the SIS is used to develop institution specific predictive models for students providing prediction of students' performance in a course from day one (i.e. earlier than many other initiatives). Data

from the LMS, grades and engagement data are used to further fine-tune the predictions during the teaching periods. The prediction results are shared with academic advisors and administrators via a number of methods including apps for action analytics to understand risk and success factors and target and test interventions. Tools for interventions such as personalised emails are used. The main form of interventions includes email and in some cases phone calls from advisors are supplemented for at-risk students. The case studies have been deployed over multiple teaching periods with increasing student numbers and courses in the three institutions. It has been shown that the test groups show student success over the control group with statistically significant results after multiple iterations of deployment over teaching periods. Several insights from the case studies are shared. Milliron et al. (2014) argue for developing institution-specific predictive models – that is, there is no one-size-fits-all predictive model—using institutional data sources such as SIS and LMS data with other relevant data sources adding to the understanding of student progress and success factors. This is different to the approach from Jayaprakash et al. (2014) where they investigated using a single predictive model across multiple institutions. Other findings specify iterative nature for interventions (trying and testing action analytics – “no silver bullets” to interventions) and bringing the insights (prediction results) to the right stakeholders for action. The authors recommend getting the four rights to LA interventions in institutions: (a) building the right infrastructure to (b) bring the right data to (c) the right people in (d) the right way. The right way is the most challenging “because it includes how we visualize data, operationalize interventions and outreach, choose modalities, provide real-time feedback, and test the timing of interventions and outreach” (Milliron et al., 2014, p. 81). Similar to CS (Arnold & Pistilli, 2012), this study involves a large number of students (over 5000 students in some test groups) while also evaluating it across multiple institutions.

In another study, Burgos et al. (2018) applied logistic regression to predict students at risk of dropping out in the course early based on students’ assessed activities. The predictive model was deployed in five courses (104 students) to detect potential dropout students. They created an intervention tutoring plan enabling instructors and tutors to intervene at different weeks of the term (weeks 4, 7 and 10) to advise students who are at risk of dropping out of the course. The intervention by instructors/tutors contacting at-risk students via email and phone resulted in a reduction of 14% in the dropout rate compared to previous cohorts.

Choi et al. (2018) identified at-risk students in an undergraduate business quantitative methods course by creating predictive models using clicker responses from formative assessments, student demographic data and summative assessments, which are easily accessible to instructors. The predictive model uses hierarchical logistic regression (LG) and hierarchical linear regression (LR) models, which was effective in identifying at-risk students at an early stage. A systematic proactive advising approach called “intrusive advising” (Earl, 1988; Varney, 2007, 2012) was used as the intervention strategy for at-risk students. Their result shows that the intervention success rate increases correspondingly with the number of interventions, and the intervention effects on peer groups are far more successful than on individual students. Overall, the students’ pass rate in the intervention group was 7% higher than that for the whole course.

A similar approach to CS is used by a team of academics at the Universitat Oberta de Catalunya (UOC), a fully online university, which developed an Early Warning System (EWS) to predict at-risk students using data from the institutional data mart (Baneres et al., 2019). The system employs Green-Amber-Red signals on dashboards to inform students of their progress and minimum marks required to pass the course, while also providing teachers with a more detailed dashboard. A number of intervention studies with different courses are conducted at UOC (Baneres et al., 2019, 2020; Guerrero-Roldán et al., 2021). The EWS has shown benefits including early detection of potential at-risk students, better guidance of students with visualisation dashboards and feedback, increment of the interaction with at-risk students (Baneres et al., 2019), and increase in performance and reduction in the dropout rate (Baneres et al., 2020, 2023; Guerrero-Roldán et al., 2021).

Borrella et al. (2022) evaluated different types of interventions to minimise dropouts in Massive Open Online Courses (MOOCs), a prevalent issue in the context of MOOC learning environments. According to Lee and Choi (2011), dropout factors can be classified into three categories: student-related factors (personal), course/program-related factors (institutional), and environmental factors (external). Personal factors include demographic background (e.g. gender), individual traits (e.g. determination, self efficacy, motivation), and academic background (e.g. digital skills). Institutional factors are highly dependent on the pedagogical approach of the course (e.g. design elements such as content format, modality and typology of activities, assessment etc.). External factors are usually unexpected events in students' lives (e.g. financial issues, family and work commitments, etc.). Borrella et al. (2022) suggest that most institutional factors (course content, support, communication, etc.) are within the course team's control and can be used as levers to influence students' experience and personal factors and ultimately their dropout decision. Borrella et al. (2022) evaluated four different interventions (A, B, C and D) to minimise dropouts in MOOCs. Interventions A and B used ML predictive models to identify students at-risk of dropping out and conducted interventions. Intervention A provided email communication intending to increase motivation before an important assessment for students at risk of dropping out. The email was drafted following the ARCS (Keller, 1987) model. Intervention B provided preparation materials and study guidelines before exams for students at risk of dropping out. A control and treatment group was used to evaluate the impact of interventions. It was found that there was no statistically significant impact of dropout rate due to these interventions. In interventions C and D, at-risk students were identified using data analysis to identify targets of the interventions. In intervention C, the assessments gradually increased in difficulty. The hypothesis was that facing a higher level of difficulty in assessments at the beginning of the course can lead to higher dropout rates, due to perceived difficulty, even if the student is able to pass the course. In intervention D, the most difficult sections of the course were identified and re-designed with scaffolding to gradually enable students to improve understanding of the content. Evaluating interventions C and D showed statistically significant reduction in dropout rates. Their study was able to provide evidence that in the MOOC context, ad-hoc interventions via personalised emails and extra materials did not impact the dropout rate. However, taking actions, identifying factors that affect a negative outcome in courses (e.g. dropouts) and

addressing them using a didactic scaffolding learning approach for topics students perceive difficult, and gradually increasing difficulty in assessments, showed promise in reducing the dropout rate in MOOCs.

There are a number of other LAI studies (Larrabee S nderlund et al., 2019; Wong & Li, 2020) that have shown positive impacts (Cambruzzi et al., 2015; Corrigan et al., 2015; Dodge et al., 2015; Rahal & Zainuba, 2016; Dawson et al., 2017; Lu et al., 2018; Espinoza & Genna, 2021; and Wang et al., 2022). We observe that even though LAIs have shown promise of positive outcomes, LAI's uptake/adoption is still in its infancy.

Gaps and Challenges in LAIs: Deriving the Research Questions

There has been evidence of positive impacts of LAIs since 2012 (Arnold & Pistilli, 2012). However, we do not observe widespread adoption or uptake of LAIs in mainstream higher educational settings. There are a number of challenges that inhibit the adoption of LAIs. Firstly, institutions need to invest in infrastructure for LAIs prior to academics and institutions being able to pilot LAIs – such as developing institution-specific student performance predictive models by extracting data from various IT systems such as SIS, LMS and others; and dashboards/apps to disseminate information to stakeholders for decision-making and tools for interventions. Educational institutions find technological barriers to scaling up LA interventions (Lonn et al., 2015).

In LAIs, the specific interventions are under the discretion of the stakeholders (mainly the educators). There is no guarantee of success for an intervention. Interventions are considered the last phase in the Learning Analytics Cycle (Clow, 2013) and also the most challenging (Rienties et al., 2017; B. T.-m. Wong & Li, 2020). Interventions are context-sensitive and it is a challenge to identify the “optimal” interventions—what works and what does not and under which conditions (Rienties et al., 2017). The uncertain impact of different types of intervention on learners' attitude, behaviour and cognition is another challenge (Rienties et al., 2017). At-risk students may be weak at interpreting the learning analytics data and taking actions, which requires strong metacognitive skills and self-regulation (Wise, 2014). Academics may not be aware of LAIs and also skilled in developing effective interventions. Milliron et al. (2014) assert that there is no “silver bullet” to interventions and that it is an iterative process. In the literature, we observe LAI studies predominantly provide students' risk levels to academics and leave the targeted interventions under the discretion of stakeholders (academics, administrators and even students for self-regulated learning) (Arnold & Pistilli, 2012; Milliron et al., 2014 and others) who are aware of the learning context and able to make decisions to implement specific interventions. There are attempts to guide interventions with pre-determined templates and messages (Jayaprakash et al., 2014 and others). Given the complexities due to the context of learning, students' diversity, and others, a “one-size-fits-all” intervention does not work. Rather, targeted interventions need to be personalised based on the learning context, students, and other factors. Academics may not necessarily be skilled and experienced in

providing effective interventions. It is important to provide assistance and guidance to stakeholders in providing targeted and effective interventions.

LAI studies measure the impact of an intervention comparing between an intervention cohort (experimental group) and the non-intervention cohort (control group) (e.g. using metrics such as pass, fail, dropout rates, etc.). However, there is little evidence that after the first intervention in a course iteration, applying interventions in future iterations of the course will show improvement compared to the previous course iteration. Is there a possibility to have a model that improves educational outcomes for future iterations of a course after LAIs are applied or is LAI only impactful for a single course iteration?

Given the fact that there is no standardised approach to interventions, providing effective interventions is considered a challenge and the fact that the success of interventions is not guaranteed and even limited to a single iteration of a course may further disincentivise institutions to invest in LAI infrastructure and educators to pilot LAI strategies.

We derive the following research questions to address the gaps discussed above:

- RQ1: Can a framework for LAI infrastructure be provided without the need for institutional-level investment to encourage the uptake in LAIs?
- RQ2: Can assistance and guidance be provided to develop effective interventions in LAIs?
- RQ3: Can a model for continuous improvement in educational outcomes in LAIs be proposed?

If a framework is to be developed without the need for institutional-level investment in LAI infrastructure, which can be accessed conveniently by the relevant stakeholders (i.e. academics), then academics can pilot LAIs in their courses. This approach encourages the uptake of LAIs and stakeholders piloting LAIs in their courses, addressing RQ1.

It is generally not conceivable to have a single intervention strategy that can be applied in all situations (i.e. one-size-fits-all model). Thus, we expect stakeholders (academics and administrators) to develop targeted effective interventions for predicted at-risk students. It would be ideal for an LAI framework to provide assistance and guidance (in terms of relevant information and approaches) for academics and administrators to develop effective interventions, addressing RQ2.

In the literature, impact of LAI studies is determined by comparing the impact of an intervention with a control group that was not exposed to interventions. Can we expect further improvements in educational outcomes using LAIs in course iterations after LAIs are applied in a previous course iteration? That is, a model for continuous improvement be possible using LAIs, addressing RQ3.

The above research questions aim to address the gaps and challenges identified in the uptake and adoption of LAIs in tertiary education and are the focus in this study. The next section discusses the study's design to address these research questions.

Study Design

To address RQ1 and RQ2, this paper presents the SPPA framework. The conceptual vision for the SPPA was presented in our earlier work (Alalawi et al., 2021a). SPPA provides a self-service model for academics to pilot LAIs in their courses without the need for institutional investment in LAI infrastructure that integrates data from diverse IT systems to develop student performance predictive models.

SPPA allows academics to develop course-specific predictive models based on historical continuous assessment data for the course. This approach avoids the need for integrating diverse data sets from an institution's IT systems. Our previous work has shown the ability to develop high-performing course-specific student performance predictive models using continuous assessment data (Alalawi et al., 2021b). The main limitation of using only continuous assessment data is that student performance prediction can be used and interventions applied only after the students submit a continuous assessment task in their course. Of course, it makes sense to provide interventions after at least the first continuous assessment task is due where students had the chance to submit assessable work demonstrating their level of learning in the course so far (Guerrero-Roldán et al., 2021). In SPPA, it is encouraged to have continuous assessments as early as possible, when designing the course, to gauge students' level of learning early on. SPPA was first evaluated in a computing course in a tertiary-level education institution (Alalawi et al., 2024). This study extends the evaluation using a number of courses acquiring further evidence to its approach.

SPPA provides a number of LA tools to facilitate interventions such as predictive models, dashboards, and integrates with tools for personalised group email messaging. Similar to other LAI approaches (Arnold & Pistilli, 2012 and others), SPPA leaves the specific intervention strategy to the discretion of the academic, who is aware of the learning context and has authority to make decisions on specific interventions, with the aim to assist them to provide effective interventions. To address RQ2, SPPA considers the entire lifecycle of the course, not only the interventions during course delivery, to enable effective interventions. Pedagogy principles and learning theories, such as Bigg's Constructive Alignment (CA) (Biggs, 2014) and Hattie and Timperley's (2007) model for effective feedback are integrated to SPPA. The course is designed/re-designed based on CA where course learning outcomes are mapped to teaching and learning activities (TLAs) and assessment tasks (ATs). This information is mapped to a CA Mapping Model and used to identify students' knowledge gaps and provide personalised study/revision plans for use during targeted interventions and also during course evaluation phase. Further details of SPPA are provided in the next section.

To answer RQ1 and RQ2, we deploy SPPA in six courses in a tertiary education environment and collect and analyse the data. We pose and answer the following sub research questions for RQ1 and RQ2:

- Sub RQ1: Can we develop high-performing predictive models using continuous assessment data?
- Sub RQ2: Can SPPA facilitate effective interventions?
- Sub RQ3: Are there any observable impacts of SPPA on student cohorts?

- Sub RQ4: What are the views and acceptance of educators who used SPPA?

To address RQ3, we pose the following sub research question:

- Sub RQ5: Does using SPPA in multiple iterations of the course result in continuous improvement?

It is important for SPPA to accurately predict students' performance from continuous assessments as this is crucial to identify at-risk students for targeted interventions. To answer this question, Sub RQ1 is proposed and addressed. We will use a variety of metrics (i.e. accuracy, F-measure, recall, and precision) to measure the performance of the predictive models generated.

Sub RQ2 aims to identify any observable impact of targeted interventions. To measure the impact of targeted interventions, a quasi-experiment with control and experimental groups are selected, comparing their performance metrics (i.e. pass rates, fail rates, and final grades). We apply predictive models in a cohort of students where SPPA was not used to identify at-risk students (control group) and compare with the students identified as at risk and targeted for interventions (experimental group). To identify any impacts of interventions, statistical tests are performed on a number of metrics including pass rates, fail rates, and average grades, between these two groups. To identify any statistically significant differences between the two groups, a chi-square (χ^2) contingency test (Rao & Scott, 1984) is used to compare the pass/fail rates. Fisher's exact test is used as an alternative to the chi-square test, in situations of small sample sizes, typically when the sample size falls below 10 (Campbell, 2007). Additionally, independent t-tests are conducted to examine the differences in final marks between the experimental and control groups. By doing so, we aim to determine an observable impact using SPPA in academic-led interventions for predicted at-risk-of-failure students.

It is possible that SPPA may have had an impact on the overall cohort (not only the targeted intervention students). To evaluate the overall impact of the SPPA framework, Sub RQ3 investigates whether the framework has any noticeable (i.e. statistically significant) impact on the performance of the entire student cohort. To achieve this, a control group, who was not exposed to SPPA, and an experimental group, who was exposed to SPPA, is established with similar characteristics, and their academic performance metrics (i.e. pass rates, fail rates, and mean final marks) are compared using statistical tests (i.e. a chi-square (χ^2) contingency test to compare the pass/fail rates and independent t-tests to examine the differences in final marks between the experimental and control groups).

Propensity Score Matching (PSM) (Austin, 2011; Rosenbaum & Rubin, 1983) is used to obtain a similar cohort as the experimental group. PSM is used in situations when randomised controlled trials are not possible. In this study, PSM was employed for each course to match a control group (students who were not subjected to SPPA) containing students from a 2021 cohort of the course and the experimental group (students who were subjected to SPPA) containing students from a 2022 cohort. Four baseline student characteristics, gender, program entry score, age, and citizenship (whether the student is domestic or international), were used as parameters in

PSM. The program entry score (UAC rank between 0 and 99) is the percentile score used for Australian University admissions representing how well Year 12 (Secondary School) students have performed in their examinations compared to their peers and only applies to domestic students who studied Secondary Schools in Australia and does not consider other pathways to university entry. The propensity score of students is calculated using logistic regression, with a binary outcome of whether the student was enrolled in a course where SPPA was deployed and independent attributes being the four baseline characteristics. PSM is carried out as a one-to-one match using the nearest neighbour method, which locates the closest match based on the distance between propensity scores. Therefore, a control match is chosen as the individuals with the closest propensity score for each treatment subject. The PSM produced two equal-sized treatment and control groups for each course evaluated.

Academics are the primary users of SPPA, who will decide to use SPPA to pilot LAIs in their courses. It is important that their views are considered and evaluated for the uptake of SPPA and its approach which is focused on Sub RQ4. We conduct interviews with academics who used SPPA to collect qualitative data and analysed them. A thematic analysis was undertaken to interpret and analyse interview data. Thematic analysis was chosen for its versatility and ability to identify emerging themes across various epistemological and theoretical approaches (Braun & Clarke, 2006). The interviews were transcribed, and the content was classified into themes and concepts (Lune & Berg, 2021).

Sub RQ5 aims to identify if repeated use of SPPA with academic's interventions provides evidence of continuous improvement. We iteratively apply SPPA in two course iterations and evaluate any observable impacts in pass/fail rates and final grades using independent t-tests.

The next section presents SPPA framework.

SPPA

The SPPA framework considers the entire life cycle of a course, including Course Design, Course Delivery, and Course Evaluation phases (see Fig. 1). During the Course Design phase, the curriculum and course content are developed/re-designed. Sound pedagogy principles/approaches are used in re-designing/developing course content, assessments and teaching/learning activities. Additionally, predictive models are developed, and their performance is evaluated during this phase. The Course Delivery phase involves the actual delivery of the course during a teaching period, and the predictive models are deployed to identify at-risk students and provide appropriate interventions. The Course Evaluation phase reflects on the results of the previous two phases and aims to identify impacts of interventions as well as issues or areas for further improvement. Reflection on the course and further optimization of the course for the next iteration is considered here. The framework facilitates continuous monitoring, improving, and evaluation cycles. SPPA is implemented as a web application and is accessible by course instructors (also called academics in this paper) and students via a web browser.

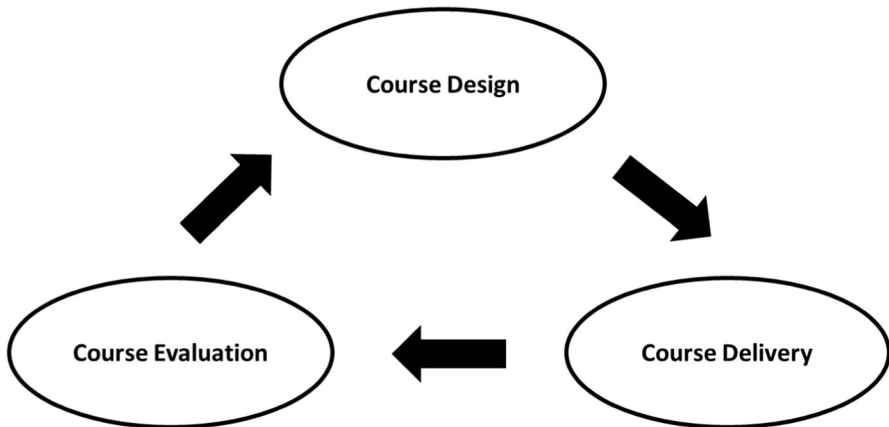


Fig. 1 Course Life Cycle

Course Design Phase

The Course Design phase of the SPPA framework entails the design of course and interventions utilizing sound pedagogy principles/approaches and the development of the predictive models.

Predictive Models

In SPPA, the historical assessment data from the course is utilised in developing the predictive models (Alalawi et al., 2021b). Historical assessment data sets for a course are the most accessible for course instructors to obtain without the need to access institution's data sources. The reliance on big data at the institutional level was reported as a challenge for successful implementation of LA (Fang & Shewmaker, 2016).

SPPA creates course-specific predictive models instead of institutional models—taking into consideration the instructional conditions, which vary across disciplines, courses, and instructors' preferences, instead of targeting a one-size-fits-all predictive model (Choi et al., 2018; Gašević et al., 2016).

SPPA implements five well-known machine learning (ML) algorithms (Logistic Regression, Support Vector Machine, Decision Tree, k-Nearest Neighbours, and Naive Bayes) to generate student performance predictive models. The ML-based predictive models are generated using the course's historical continuous assessment data for each assessment task and evaluated using precision, recall, accuracy, and F-measure as evaluation metrics. These evaluation metrics rely on parameters such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) that are calculated based on the confusion matrix. The performance evaluation metrics are defined as follows:

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F - Measure = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \quad (4)$$

During the creation of ML predictive models, SPPA employs a 70–30 data split, allocating 70% for training and 30% for testing. SPPA further fine-tunes the models by utilizing fivefold cross-validation to optimise hyperparameters for each ML algorithm. Once SPPA has determined these optimal hyperparameters, it proceeds to construct the predictive models. This ML model creation process is automated and entirely handled by SPPA.

The predictive models, generated by SPPA, come in two forms: binary predictive models (distinguishing between likely to pass and likely to fail) and multi-classification predictive models (categorizing as likely to fail, borderline, or likely to pass) (Alalawi et al., 2021b). SPPA develops these models using historical continuous assessment scores obtained during the course design phase, prior to course delivery.

SPPA selects the best model for student performance prediction based on the highest accuracy followed by F-measure, recall, and precision by default. Given the simplicity of the data sets and format required for creating predictive models in SPPA, academics can conveniently load historical student assessment data for the course to generate the predictive models. This approach avoids the need for technical and human resources, efforts in data processing, which is estimated to be as high as 85% of the cost of LA implementation (Bienkowski et al., 2012). Figure 2a shows the format course instructors use to upload the data and Fig. 2b shows the results of the predictive models' metrics after each continuous assessment. Note that the prediction results improve over time as the more continuous assessment data are available for predictive models towards the latter part of the teaching period.

Course Design using Constructive Alignment

SPPA emphasizes using sound pedagogical principles and approaches in course design and interventions. SPPA utilises Constructive Alignment (CA) (Biggs, 2014) for course design and creating the course's CA Mapping Model. In Constructive Alignment, the intended outcomes that the student needs to learn or demonstrate (termed, Intended Learning Outcomes – ILOs) are clearly specified. In a course, these outcomes are termed Course Learning Outcomes (CLOs). The assessment tasks (ATs) are designed to evaluate the level of achievement by students with respect to the course learning outcomes. The teaching and learning activities (TLAs) are designed

	A	B	C	D	E	F	G
1	PracticalTest 1	Assignment 1	Assignment 2	PracticalTest 2	Final Exam	Total (out of 100)	Grade
2	50	20	25	50	80		
3	20	10	10	20	40		
4	47.5	14	25	48.75	69.6	90	HD
5	42.5	19	17.5	16.25	46	63	P
6	37.5	18	15	43.75	43.6	69	C
7	40	12	12.5	43.75	42.4	66	C
8	43.75	18	22.5	48.125	56.4	83	D

(a.)

Course Design:

- Create Course
- Create CLO
- Create Assessment for Course
- Create TLA
- Show Revision Plan
- Update CLO AT and TLA
- Update Mapping
- Show Mapping Model

Prediction Model:

- Upload Dataset
- Create ML Models
- Perform Prediction & Upload Students
- Mark
- Update Existing Models
- Delete Course and Model

Dashboards:

- Mail Merge Data
- Academic Dashboard
- Student Dashboard
- Upload Intervention Data
- Final Course Report

Thank you, your models are created successfully

Please note that: [805] students records have been used for creating ML models after data cleaning and preprocessing. The criteria preprocessing and cleaning the dataset is: (1) students who dropout the course early in the semester, (2) students who did not attend received (F) grade.

Five ML Models have been created for [REDACTED] course:

Logistic Regression Binary model performance:

	Practical Test 1	Assignment 1	Practical Test 2	Assignment 2
Accuracy	0.887	0.906	0.934	0.904
F-Measure	0.937	0.947	0.982	0.979
Recall	0.987	0.973	0.983	0.977
Precision	0.893	0.923	0.943	0.881

Logistic Regression Multiclass model performance:

	Practical Test 1	Assignment 1	Practical Test 2	Assignment 2
Accuracy	0.805	0.829	0.884	0.899
F-Measure	0.75	0.801	0.873	0.895
Recall	0.805	0.829	0.884	0.899
Precision	0.718	0.794	0.878	0.895

(b.)

Fig. 2 a Format to load data; b SPPA displays metrics for predictive models generated

to optimize student learning to achieve the CLOs. In essence, CLOs, ATs and TLAs are aligned in CA.

In SPPA, instructors are encouraged to use CA to design/re-design their courses and create a CA Mapping Model that maps the course's CLOs with the ATs and TLAs.

CA Mapping Model

Instructors specify the course's CLOs and map them to ATs and TLAs in the CA Mapping Model. This activity guides to develop well designed courses where the course's learning goals are sufficiently assessed in the ATs and that TLAs are designed to help students achieve levels of learning as specified by the CLOs. Academics can specify the mapping at a coarse or fine granularity, and we expect that the mapping model to be fine-tuned and mapped to fine granularly over multiple iterations of the course. The mapping model is used in personalised feedback and course evaluations, as will be discussed below.

In the example (see Table 1), the Course Learning Outcome 1 (CLO1) is divided into two granular levels (CLO1a and CLO1b). Each CLO is mapped to an Assessment

Table 1 Example CA Mapping Model

Course: Course A	
Course Learning Outcomes	Teaching and Learning Activities
CLO1: Design and configure peer-to-peer networks to share resources	Assessment Tasks <ul style="list-style-type: none"> • Practical Test 1-Design Task • Final Exam • Practical Test 1-Practical Task
CLO1a: Design a peer-to-peer networks to share resources	
CLO1b: Configure a peer-to-peer networks to share resources	

Task(s) which evaluates the level of learning for the CLO. Of course, the ATs can also be specified in fine-granular – such as Final Exam’s questions 2, 5, and 6 can be mapped to CLO 1a instead of coarse grain mapping of “Final Exam”. Similarly, TLAs can be specified at a coarse or fine granularity. For instance, Lectures 1–4 has coarse granularity while the class exercise—“Network design exercise in class (Absolute Cleaning)” is a fine-granular TLA. The mapping model having finer granularity will be able to provide more detailed information when providing feedback and during course evaluations. The mapping model is expected to be revised to be fine-granular over multiple iterations of the course by the instructors as they realise their value upon using this feature.

Design of CLOs, ATs and TLAs

Taxonomies such as the Structure of the Observed Learning Outcome (SOLO) (J. B. Biggs & Collis, 2014) and Bloom’s Taxonomy (Anderson & Bloom, 2001) can be used to specify CLOs and provide guidance on the levels of student learning that need to be evaluated. The design of effective teaching and learning activities and assessment tasks for the course’s context can incorporate pedagogy theories and principles in education field, such as scaffolded learning (Maybin et al., 1992; Vygotsky & Cole, 1978), formative assessments (Black & Wiliam, 1998), design TLAs following principles of Kolb’s experiential learning (Kolb, 1984), project-based learning (Krajcik & Blumenfeld, 2006), personalisation based on different learning styles (Fasihuddin et al., 2017) and others. Academics are encouraged to use best-practice approaches and techniques in designing CLOs, ATs and TLAs considering the learning context.

Course Delivery Phase

During the Course Delivery phase, students at-risk of failing are identified and interventions take place. In SPPA, there are several LA tools available to academics.

Predicting At-Risk students

In SPPA, instructors can run the predictive models after each continuous assessment. In SPPA, instructors upload the latest continuous assessment data of the current cohort to SPPA’s prediction module in a particular format (see Fig. 3a) which provides predicted results for at-risk students (see Fig. 3b, c). This information is also available via the Academic dashboard (see example in Fig. 4). Students who are predicted as “likely to fail” in the binary predictive model as well as students predicted to be “likely to fail” or are in the “borderline” category in the multi-class predictive model are classified as at-risk in the Academic dashboard.

Providing Effective Feedback

Once students at-risk of failure has been identified, instructors have the option to provide academic-led interventions. Students are provided with feedback during

	A	B	C	D	E
1	Student Name	Student ID	Practical Test 1	Assignment 1	Assignment 2
2			50	20	25
3			20	10	10
4	Student 1	1234567	47.5	14	25
5	Student 2	1234568	42.5	19	17.5
6	Student 3	1234569	37.5	18	15
7	Student 4	1234570	40	12	12.5
8	Student 5	1234571	43.75	18	22.5
9	Student 6	1234572	40	18	19.75
10	Student 7	1234573	38.75	16	13.5

Assessment's name

Assessment's marked out of

Assessment's weight (% out of 100)

Unconverted marks

(a.)

Results of the SVM Binary model with 97.3% accuracy:

Student Name	Student ID	Prediction Result (Pass [1] /Fail [0])
Student 1	123	1
Student 2	456	1
Student 3	789	0
Student 4	147	1

Model used for prediction (binary classification)

Prediction results for binary classification

(b.)

Results of the LR Multiclass model with 94.1% accuracy:

Student Name	Student ID	Prediction Result (Pass/Borderline/At-Risk)
Student 1	123	borderline
Student 2	456	pass
Student 3	789	at_risk
Student 4	147	pass

Model used for prediction (multiclass classification)

Prediction results for multiclass classification

(c.)

Fig. 3 **a** Format of data to be uploaded to perform prediction; **b** Sample results of binary predictive models; **c** Sample results of multi-class predictive models

interventions. According to Hattie and Timperley (2007, p. 81) “Feedback is one of the most powerful influences on learning achievement...” but “the type of feedback and the way it is given can be differentially effective”. According to Hattie and Timperley (2007), effective feedback needs to answer three main questions:

- Where am I going? (What are the goals?),
- How am I going? (What progress is being made toward the goal?), and
- Where to next? (What activities need to be undertaken to make better progress?)

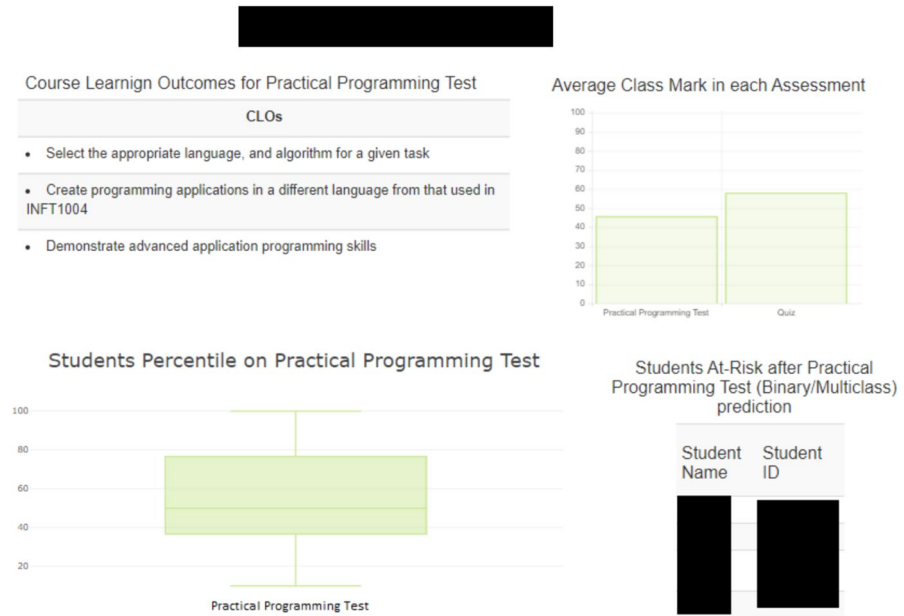


Fig. 4 Academic dashboard showing students at-risk of failure

SPPA encourages and facilitates effective feedback based on the Hattie and Timperley (2007) model for effective feedback. In SPPA, CA Mapping Tool maps ATs to CLOs explicitly. Thus, the ATs have specific and expected learning “goals” and levels of achievement clearly specified which can answer the question “Where am I going? (What are the goals?)”. Next, marks for assessment and instructor’s feedback for each section of the AT addresses the question—“How am I going?”. Finally, if there are any knowledge gaps between what is expected and areas for improvement, these can be identified as the CA Mapping Model maps ATs to TLAs. If fine-grained mapping is available, reviewing the specific TLAs that map to sections of the AT that the student has performed poorly can provide specific areas for students to revise and address the question “What activities need to be undertaken to make better progress?”. Instructors can also provide supplementary content and guidance to areas where students struggle during the interventions as well as provide advice on preparing for future assessment tasks. Figure 5 provides a sample student dashboard in SPPA which provides feedback for an assessment task to a student based on Hattie and Timperley (2007) model for effective feedback. It’s important to note that the revision plan (as shown in Fig. 5) is automatically generated based on the CA Mapping Model. The instructor/tutor comments, are added manually by the academic, typically referencing the LMS’s submission where detailed feedback is provided.

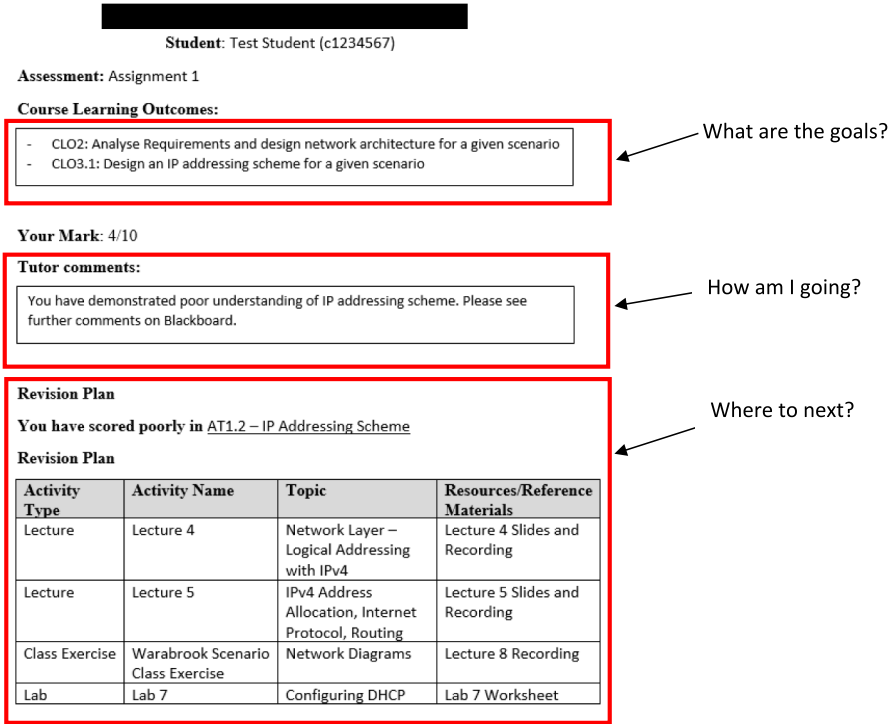


Fig. 5 A sample student dashboard in SPPA

Tools for Personalised Interventions

SPPA allows instructors to provide personalised feedback. Pardo et al. (2018)’s OnTask provides an architecture where data from multiple sources are integrated to create a single large student information table. Next, rules are used by the instructors to select groups of students for personalised email interventions. SPPA incorporates this idea to enable personalised email interventions. Data is collected from the class list, assessment data and prediction results. These data sets are integrated to a single Student Table (see Fig. 6a, b). In SPPA, the student table is imported to a spreadsheet and filters are used to select student groups by academics for personalised interventions and emails are generated using Mail Merge tool in Microsoft Word (see Fig. 6c, d). Academics are encouraged to follow the model by Hattie and Timperley (2007) for effective feedback in generating personalised emails and can attach reports from student dashboard (e.g. personalised revision plans). The academics draft the intervention emails with fields added in Mail Merge to allow personalisation (see example in Fig. 6d). We expect academics to save these email templates and use them in future iterations of the course during interventions.

Fig. 6 **a** Workflow for personalised email interventions; **b** A student big table integrating student data, assessment results and prediction results; **c** Filters used in Mail Merge to select groups of students; **d** A sample Mail Merge email with personalisation using inserted fields (such as student name, assessment scores etc.)

Course Evaluation Phase

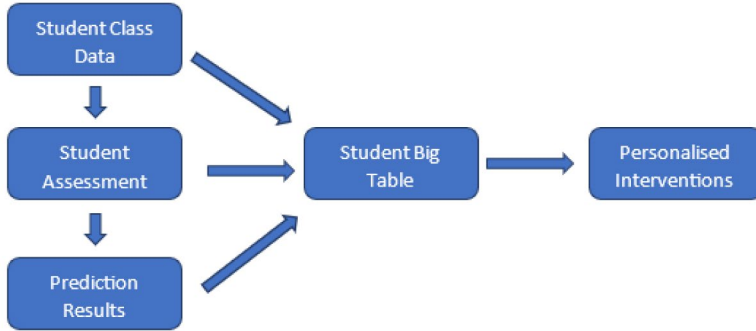
The aim of the course evaluation phase is to evaluate the course iteration. SPPA provides a number of visualisations and reports including the *Overall Course Performance* report (see example on Fig. 7) which compares pass rates, fail rates, withdrawal rates and average final grades with previous offerings of the course. *The Course Assessment Performance* report (see example on Fig. 8) presents average grade for each assessment and compares with current and previous offerings. An assessment with a low average mark may point to an area for improvement. If a particular assessment is performing poorly, this typically requires an investigation and review of the AT and TLAs mapped to the AT which can lead to identify areas for improvement for next iteration of the course. *The Intervention Performance report* (see example on Fig. 9) evaluates the effectiveness of the interventions. The report presents students who were subject to interventions, their final grade and their predicted grade. Descriptive analytics such as percentage of students passing, failing, withdrawing in the at-risk group is provided. These reports and visualisations along with other sources of course evaluations (such as student feedback on course) can assist instructors to reflect and identify issues and areas for improvement for future iterations of the course.

Overall, SPPA allows academics to pilot LA interventions in their courses conveniently and flexibly. Academics only require access to historical continuous assessment data for their courses to develop course-specific predictive models. Academics access SPPA's functionality via a web browser to pilot LA interventions without the need for any configuration or additional institution investment in LA intervention infrastructure. Also, SPPA's source code is available as an open-source project¹ for further research and development and educational institutions can also host SPPA on their own IT infrastructure.

Evaluation

Relevant ethics approval was obtained by the University's ethics committee prior to undertaking this evaluation. A total of six courses were recruited for the study. Four academics were involved in teaching the six courses. The academics were introduced to SPPA and its features. Academics were given the freedom to perform predictions and targeted interventions. The details of courses used in this study's evaluation are provided in Table 2.

¹ <https://github.com/khalidalalawi/Student-Performance-Predictive-and-Action-SPPA-git>



(a.)

Student ID	Student Name	Practical Test 1 (Weight: 20%) (Marked out of:20)	Prediction results (binary)	Prediction results (multi-class)	Assignment 1 (Weight: 10%) (Marked out of:10)	Prediction results (binary)	Prediction results (multi-class)
S001234	John Doe	8	at-risk	at-risk	4	at-risk	at-risk
S002453	Mary Once	18	pass	pass	9	pass	pass
S238894	Kate None	12	at-risk	borderline	6	pass	borderline

(b.)

(c.)

Dear «Student_Name»,

We have concern with your progress in the ABC course. Your continuous assessments scores (shown below) are not encouraging.

Assessment	Score
Practical Test 1	«Practical_Test_1»
Assignment 1	«Assignment_1»

I would strongly encourage to contact me to discuss your Practical Test 1 and Assignment 1 scores as well as discuss ways to improve your performance by taking the practice test and class exercise before the next Assignment. Note that if you have valid reasons as discussed in class, you may be eligible to submit an alternate assessment task to improve your continuous assessment scores

Kind regards

Instructor/Tutor

(d.)

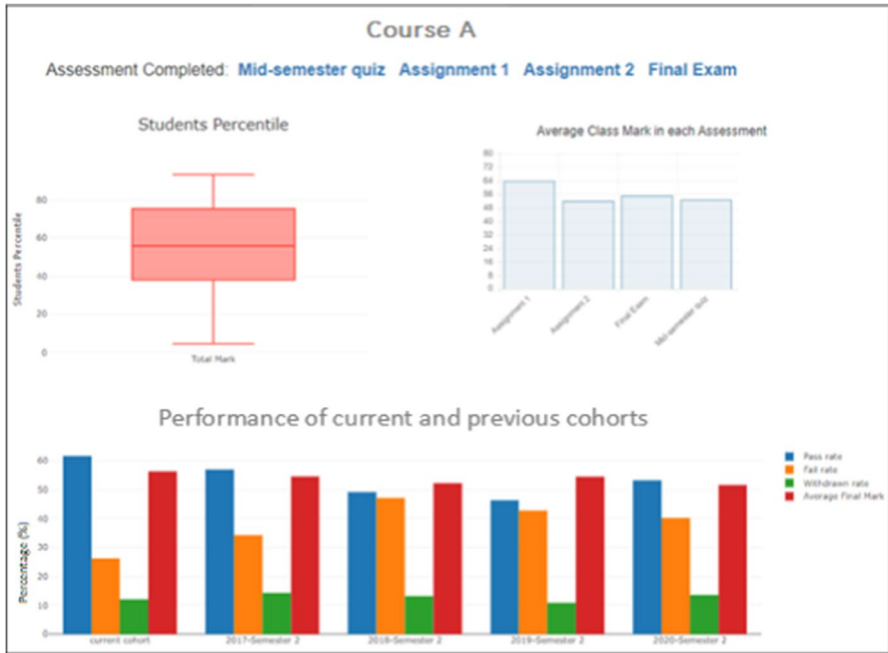


Fig. 7 Overall course performance report

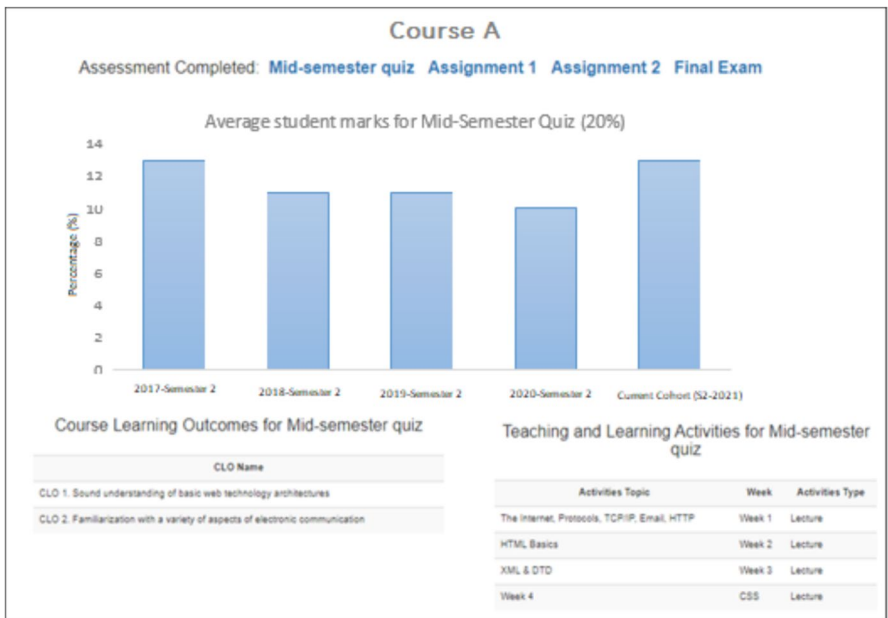


Fig. 8 Course assessment performance report

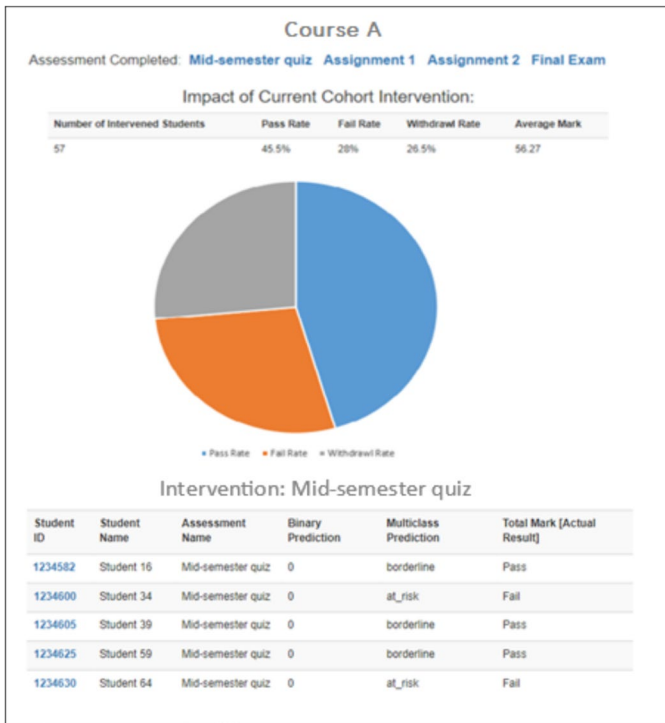


Fig. 9 Intervention performance report

The courses used in the study are briefly described below:

- **Course A:** introduces students to contemporary astronomy’s fundamental concepts, procedures, and methods. It supports students’ use of analytical thinking and problem-solving techniques to address Astronomy-related issues. The course runs for 13 weeks, and students attend 12×2 -hour lectures, 4×3 -hour laboratory practical sessions, and 12×1 -hour tutorials during the semester. Students are expected to prepare for every practical session by completing pre-laboratory activities online in the LMS—Canvas. They are also expected to reinforce their learning in the lectures by completing weekly online quizzes on Canvas. The course comprises five assessments whose marks contribute to the final course grade: a mid-term test (10%), ten weekly quizzes (10%), four assignments (20%), lab activities (20%), and final exam (40%). Anonymized data from 92 students from 2021 and 2022 over the course were collected for analysis. The course design and content were the same across the two cohorts, and data from 2021 cohorts ($n_{2021} = 35$) were used for comparison against the 2022 cohort ($n_{2022} = 57$) who were exposed to SPPA. Historical academic data from 2019 to 2021 (147 student records) were used to create ML predictive models.
- **Course B:** introduces formal computing methods and the problems that students can resolve. It also introduces students to Turing machines and related models of computation, including basic constraints on what can be calculated. It includes context-free

Table 2 Courses and Assessments Details

Course (Instructor)	Assessments	Week	Weight
Course A (Instructor 1) Astronomy course	Mid-term test	7	10%
	Ten weekly quizzes	2–6 and 9–13	10%
	Five assignments	4,6,8,10,12	20%
	Four lab activities	4,7,10,13	20%
	Final exam	14/15	40%
Course B (Instructor 2) Fornal computational methods course	Assignment 1	5	15%
	Class Test	8	20%
	Assignment 2	10	20%
	Ten weekly online quizzes	2 to 12	10%
	Final exam	14/15	40%
Course C (Instructor 3) Programming course	Quiz Test	3	10%
	Practical programming test	4	20%
	Programming group assignment	11	30%
	Final exam	14/15	40%
Course D (Instructor 4) IT systems and networks course	Practical Test 1	6	20%
	Assignment 1	8	10%
	Assignment 2	12	10%
	Practical Test 2	12	20%
	Ten weekly quizzes	2 to 12	10%
	Final exam	14/15	30%
Course E (Instructor 4) Advanced data management course	Assignment 1	6	25%
	Assignment 2	8	20%
	Assignment 3	12	20%
	Final Exam	14/15	40%
Course F (Instructor 4) IT systems and networks course (online)	Practical Test 1	6	15%
	Module 1—Weekly Tasks	6	5%
	Module 2—Weekly Tasks	8	5%
	Assignment 1	8	10%
	Module 3—Weekly Tasks	10	5%
	Assignment 2	12	10%
	Practical Test 2	12	15%
	Module 4—Weekly Tasks	13	5%
	Final exam	14/15	30%

languages and grammars, non-context-free grammars, regular expressions, and finite state machines. The course runs for 13 weeks, and students attend 12×2 -hour lecture sessions and 12×2 -hour tutorials during the semester. Students are requested to prepare for every session and sit for five assessments, including weekly online quizzes. They are also expected to reinforce their learning in the lectures by completing ten weekly online quizzes in the Canvas LMS. The course comprises of five assessments whose marks contribute to the final course grade: Assignment 1 due on week

5 (15%), Class Test due on week 8 (20%), Assignment 2 due on week 10 (20%), ten weekly online quizzes from week 2 to week 12 (10%), and final exam (40%). The course has been running regularly since 2017, but some changes were made to the assessment items and weightings in 2022. Prior to 2022, the assessments weighting were as follows: Assignment 1 (15% of final grade), two Class Tests instead of one Class Test in S2, 2022: Class Tests 1 (10% of final grade) and Class Tests 2 (15% of final grade), Assignment 2 due on week 10 (15% of final grade), ten weekly online quizzes from week 2 to week 12 (5% of the final grade instead of 10% in S2, 2022), and final exam (40% of final grade). Anonymized data from 165 students from 2021 and 2022 over the course were collected for analysis, and data from 2020 and 2021 cohorts ($n_{2021}=83$) were used for comparison against the 2022 cohort ($n_{2022}=82$) who were exposed to the proposed framework. Historical academic data from 2017 to 2021 (306 student records) were used to create ML predictive models.

- **Course C:** introduces students to programming and problem-solving concepts and skills. The course aims to enhance students' understanding of program design, execution, and evaluation. It is a 13-week course that requires attendance at 12 two-hour lectures and 11 two-hour laboratory practical sessions. As part of the course activities, students are expected to prepare for each session and sit for four assessments. Course C consists of four continuous assessments that contribute to the final course grade: Quiz Test in week 4 (10%), Practical programming test in week 4 (20%), Programming group assignment in week 11 (30%), and final exam (40%). Anonymized data from 102 students ($n_{2021}=51$ and $n_{2022}=51$) who took the course in 2021 and 2022 were collected for analysis. The course has been offered regularly since 2013. Historical academic data from 2019 to 2021 (277 student records) were used to create ML predictive models.
- **Course D:** introduces students to networking and systems administration theory and practical skills for setting up client–server networks, peer-to-peer networks, and personal computers. The course runs for 13 weeks, requiring attendance at 12 two-hour lectures and 11 two-hour laboratory practical sessions. The course comprises five assessments that contribute to the final course grade: Practical Test 1 scheduled in week 6 (20%), Assignment 1 due in week 8 (10%), Assignment 2 due in week 12 (10%), Practical Test 2 scheduled in week 12 (20%), ten weekly tasks from week 2 to week 12 (10%), and final exam (30%). Anonymized data from 154 students ($n_{2021}=83$ and $n_{2022}=71$) who took the course in 2021 and 2022 were collected for analysis. The course has been offered regularly since 2010. Historical academic data from 2016 to 2021 (481 student records) were used to create ML predictive models.
- **Course E:** equips students with a comprehensive understanding of the theoretical underpinnings and practical skills necessary for working with database systems, big data, and modern data-intensive systems. This course covers an array of topics such as Advanced SQL, storage and indexing techniques, query processing and optimization methods, transaction and concurrency management, crash recovery, Object Relational Mapping, business intelligence, distributed database systems, and big data management. Spanning a duration of 12 weeks, the course entails attending 12 two-hour lectures and 12 two-hour workshops or laboratory sessions throughout the semester. The course comprises four continuous assessments, the cumulative scores of which contribute to the final course grade: Assignment 1 due in week 6 (25% of

final grade), Assignment 2 due in week 8 (20% of final grade), Assignment 3 due in week 12 (20% of final grade), and the final examination (35% of final grade). Anonymized data from 66 students ($n_{2021}=33$ and $n_{2022}=33$) who took the course in 2021 and 2022 were collected for analysis. Historical academic data from 2019 to 2021 (87 student records) were used to create ML predictive models.

- **Course F:** is an online version of Course D has the same course description, contents, and outcomes as the on-campus version. However, the course has a different assessment structure and is delivered entirely online, with one hour of online Zoom consultation per week. The course is designed for the online audience with scenario-based learning approach and problem-based learning activities embedded. The course comprises of nine assessments that contribute to the final grade: Practical Test 1 scheduled in week 6 (15%), Module 1—Weekly Tasks due in week 6 (5%), Module 2—Weekly Tasks due in week 8 (5%), Assignment 1 due in week 8 (10%), Module 3—Weekly Tasks due in week 10 (5%), Assignment 2 due in week 12 (10%), Practical Test 2 scheduled in week 12 (15%), Module 4—Weekly Tasks due in week 13 (5%), and the final exam (30%). Anonymized data from 46 students ($n_{2021}=23$ and $n_{2022}=23$) who took the course in 2021 and 2022 were collected for analysis. Historical academic data from 2019 to 2021 (95 student records) were used to create ML predictive models.

Courses A – F are used to answer sub RQ1, courses A – E are used to address sub RQ2 – RQ4 and course F is used to answer sub RQ5. PSM is used to identify control groups in sub RQ3. As discussed, PSM was used to match four baseline student characteristics: gender, program entry score, age, and citizenship (whether the student is domestic or international). Note that a value of zero is used for the program entry score for students who had taken other pathways than secondary school for entry to the program. The PSM produced two equal-sized treatment and control groups, and Table 3 depicts the size of each group in each course. PSM is carried out as a one-to-one match using the nearest neighbour method, which locates the closest match based on the distance between propensity scores. Therefore, a control match is chosen as the individuals with the closest propensity score for each treatment subject. Table 4 provides an overview comparison of the students' demographic data (the baseline characteristics) of the two groups for each course.

Results

This section describes the results of the experiments addressing the research questions (sub RQ1-sub RQ4).

Table 3 Number of cases in each group after matching using PSM

Courses	Control Group	Experimental Group
Course A	57	57
Course B	82	82
Course C	51	51
Course D	71	71
Course E	33	33

Table 4 An overview of the students' composition in each course between control and treatment groups after propensity matching

COURSES	GROUPS	# MALE (%)	# DOMESTIC STUDENTS (%)	# AGE 0-20 (%)	# AGE 21-26 (%)	# AGE 26-34 (%)	# AGE 35+ (%)	PROGRAM ENTRY SCORE (MEDIAN)
Course A	Treatment (n=57)	40 70.2%	52 91.2%	15 26.3%	29 50.9%	11 19.3%	2 3.5%	54.55
	Control (n=57)	42 73.7%	57 100.0%	17 29.8%	27 47.4%	12 21.1%	1 1.8%	52.1
Course B	Treatment (n=82)	76 92.7%	71 86.6%	23 28.0%	43 52.4%	14 17.1%	2 2.4%	64.4
	Control (n=82)	78 95.1%	69 84.1%	23 28.0%	43 52.4%	16 19.5%	0 0.0%	65.2
Course C	Treatment (n=51)	47 92.2%	43 84.3%	16 31.4%	25 49.0%	10 19.6%	0 0.0%	49.05
	Control (n=51)	48 94.1%	43 84.3%	12 23.5%	26 51.0%	13 25.5%	0 0.0%	50.85
Course D	Treatment (n=71)	64 90.1%	57 80.3%	30 42.3%	27 38.0%	11 15.5%	3 4.2%	53.375
	Control (n=71)	66 93.0%	59 83.1%	30 42.3%	27 38.0%	11 15.5%	3 4.2%	51.75
Course E	Treatment (n=33)	27 81.8%	27 81.8%	2 6.1%	22 66.7%	9 27.3%	0 0%	57.72
	Control (n=33)	29 87.9%	23 69.7%	2 6.1%	24 72.7%	7 21.2%	0 0%	55.95

Results—Sub RQ1

The Table 5 presents the range of performance metrics for binary classification after each continuous assessment, while Table 6 presents the range of performance metrics of multiclass classification after each continuous assessment. We can observe that the predictive models' performance is high on all metrics achieving a minimum 0.7 in binary classification and a minimum 0.5

Table 5 Performance range of each binary classification predictive model (i.e., LR, SVM, KNN, NB, DT) for each course (A–F)

Course	ML Algorithm	Accuracy	F-Measure	Recall	Precision
Course A (<i>n</i> = 147)	LR	0.951–0.958	0.974–0.978	0.978–1.0	0.951–0.971
	SVM	0.937–0.951	0.967–0.975	0.985–1.0	0.932–0.951
	KNN	0.951–0.965	0.975–0.982	0.993–1.0	0.951–0.965
	NB	0.858–0.908	0.919–0.951	0.874–0.956	0.957–0.979
	DT	0.88–0.944	0.938–0.974	0.911–0.978	0.949–0.972
Course B (<i>n</i> = 306)	LR	0.819–0.931	0.896–0.959	0.923–0.983	0.866–0.966
	SVM	0.815–0.859	0.898–0.923	0.966–1.0	0.839–0.86
	KNN	0.804–0.888	0.89–0.937	0.944–0.987	0.841–0.893
	NB	0.815–0.873	0.884–0.922	0.854–0.919	0.871–0.959
	DT	0.808–0.906	0.89–0.951	0.923–0.974	0.841–0.945
Course C (<i>n</i> = 277)	LR	0.744–0.916	0.842–0.942	0.919–0.956	0.754–0.936
	SVM	0.77–0.907	0.845–0.935	0.882–0.938	0.811–0.936
	KNN	0.757–0.907	0.835–0.935	0.864–0.932	0.811–0.942
	NB	0.74–0.903	0.836–0.931	0.907–0.932	0.761–0.936
	DT	0.757–0.903	0.833–0.931	0.833–0.945	0.817–0.942
Course D (<i>n</i> = 481)	LR	0.899–0.954	0.945–0.974	0.97–1.0	0.902–0.964
	SVM	0.895–0.949	0.942–0.971	0.965–0.982	0.907–0.97
	KNN	0.877–0.956	0.931–0.975	0.945–0.982	0.911–0.968
	NB	0.877–0.945	0.93–0.968	0.926–0.949	0.919–0.988
	DT	0.887–0.954	0.938–0.976	0.952–0.977	0.904–0.977
Course E (<i>n</i> = 87)	LR	0.929–0.965	0.962–0.981	0.988–1.0	0.939–0.964
	SVM	0.918–0.941	0.956–0.969	0.988–1.0	0.918–0.94
	KNN	0.918–0.929	0.956–0.963	0.988–1.0	0.928–0.929
	NB	0.918–0.953	0.953–0.975	0.935–0.988	0.963–0.976
	DT	0.859–0.941	0.918–0.968	0.898–0.961	0.964–0.975
Course F (<i>n</i> = 95)	LR	0.859–0.946	0.912–0.967	0.93–0.986	0.896–0.951
	SVM	0.848–0.891	0.903–0.936	0.903–1.0	0.84–0.93
	KNN	0.87–0.935	0.915–0.958	0.903–1.0	0.901–0.948
	NB	0.828–0.913	0.888–0.949	0.861–0.944	0.901–0.987
	DT	0.827–0.913	0.887–0.924	0.888–0.945	0.891–0.934

*KNN is (k-Nearest Neighbors), LR is (Logistic Regression), DT is (Decision Trees) and SVM is (Support Vector Machine)

***n* represents the number of historical student records used to develop the predictive models

Table 6 Performance range of each multiclass classification predictive model (i.e., LR, SVM, KNN, NB, DT) for each course (A–F)

Course	ML Algorithm	Accuracy	F-Measure	Recall	Precision
Course A (<i>n</i> = 147)	LR	0.789–0.879	0.759–0.871	0.757–0.870	0.768–0.877
	SVM	0.787–0.883	0.757–0.875	0.754–0.873	0.765–0.881
	KNN	0.786–0.880	0.758–0.872	0.754–0.870	0.764–0.879
	NB	0.767–0.863	0.749–0.855	0.740–0.853	0.758–0.861
	DT	0.745–0.840	0.720–0.832	0.708–0.830	0.730–0.838
Course B (<i>n</i> = 306)	LR	0.681–0.841	0.601–0.834	0.681–0.841	0.542–0.865
	SVM	0.674–0.721	0.609–0.649	0.674–0.721	0.564–0.707
	KNN	0.656–0.761	0.59–0.757	0.656–0.761	0.528–0.757
	NB	0.667–0.804	0.599–0.833	0.667–0.804	0.541–0.833
	DT	0.677–0.75	0.608–0.796	0.677–0.764	0.553–0.796
Course C (<i>n</i> = 277)	LR	0.696–0.813	0.683–0.791	0.676–0.77	0.667–0.793
	SVM	0.712–0.816	0.692–0.8	0.682–0.796	0.713–0.786
	KNN	0.696–0.793	0.686–0.783	0.681–0.791	0.62–0.792
	NB	0.703–0.772	0.69–0.76	0.663–0.731	0.681–0.779
	DT	0.678–0.775	0.668–0.775	0.654–0.78	0.631–0.771
Course D (<i>n</i> = 481)	LR	0.82–0.907	0.759–0.902	0.82–0.907	0.735–0.903
	SVM	0.818–0.893	0.767–0.888	0.818–0.893	0.758–0.888
	KNN	0.792–0.901	0.765–0.896	0.792–0.901	0.772–0.896
	NB	0.806–0.881	0.772–0.891	0.806–0.881	0.78–0.912
	DT	0.824–0.883	0.783–0.878	0.824–0.883	0.788–0.896
Course E (<i>n</i> = 87)	LR	0.871–0.906	0.816–0.869	0.871–0.906	0.768–0.837
	SVM	0.871–0.906	0.84–0.88	0.871–0.906	0.811–0.866
	KNN	0.882–0.918	0.864–0.897	0.882–0.918	0.851–0.886
	NB	0.847–0.871	0.827–0.859	0.847–0.871	0.801–0.858
	DT	0.835–0.894	0.879–0.885	0.882–0.894	0.889–0.899
Course F (<i>n</i> = 95)	LR	0.763–0.881	0.713–0.875	0.763–0.881	0.678–0.884
	SVM	0.785–0.805	0.728–0.768	0.785–0.805	0.71–0.775
	KNN	0.785–0.859	0.735–0.839	0.785–0.859	0.693–0.86
	NB	0.699–0.816	0.694–0.819	0.699–0.816	0.675–0.853
	DT	0.74–0.837	0.74–0.829	0.74–0.859	0.721–0.843

*KNN is (k-Nearest Neighbors), LR is (Logistic Regression), DT is (Decision Trees) and SVM is (Support Vector Machine)

***n* represents the number of historical student records used to develop the predictive models

in multiclass classification. Note that, as discussed previously, SPPA selects, by default, the predictive model with the highest accuracy first, followed by F-measure, recall and precision, respectively to identify risk-levels of students. Typically, the best models have a value above 0.86 for binary models and over 0.65 for multiclass models for all metrics.

Results—Sub RQ2

To address sub RQ2, we selected at-risk students in control group and experimental groups based on the predictive models and compared the pass rate, fail rate, and final grade of the two groups. Table 7 presents the predictive models used and their performance metrics. Note that the predictive models are high-performing on the four metrics – accuracy, F-measure, recall and precision (i.e. above 0.86 for binary predictive models and 0.65 for multi-class predictive models). These predictive models were selected by default as the best performing models by SPPA. Table 8 summarises the results. Note that dropout students are not considered in the results of Table 8.

Course A: Academic predicted student performance at two instances during the semester (in week 6 after Quiz 5 and week 10 after Quiz 9) in the experimental cohort. The experimental group identified 4 at-risk students in week 6, after Quiz 5, while 5 students were predicted at-risk in the control group. In the second prediction, after Quiz 9 in week 10, 8 students were identified as at-risk of failure in the experimental group, while 7 students in the control group. There were 11 unique students in the experimental group and 10 unique students in the control group.

Table 7 Performance metrics of predictive models at the time of intervention

Courses	Interventions	ML predictive model performance (Binary model)				ML predictive model performance (Multiclass model)			
		Accuracy	F-Measure	Recall	Precision	Accuracy	F-Measure	Recall	Precision
Course A (<i>n</i> = 147)	Intervention 1	0.965 (KNN)	0.982 (KNN)	0.993 (KNN)	0.965 (KNN)	0.908 (LR)	0.877 (LR)	0.898 (LR)	0.91 (LR)
	Intervention 2	0.958 (KNN)	0.978 (KNN)	0.993 (KNN)	0.958 (KNN)	0.916 (KNN)	0.904 (KNN)	0.913 (KNN)	0.931 (kNN)
Course B (<i>n</i> = 306)	Intervention 1	0.931 (LR)	0.959 (LR)	0.953 (LR)	0.966 (LR)	0.84 (LR)	0.836 (LR)	0.84 (LR)	0.865 (LR)
Course C (<i>n</i> = 277)	Intervention 1	0.828 (KNN)	0.878 (KNN)	0.87 (KNN)	0.89 (KNN)	0.761 (KNN)	0.757 (KNN)	0.813 (KNN)	0.791 (KNN)
	Intervention 2	0.916 (KNN)	0.942 (KNN)	0.95 (KNN)	0.936 (KNN)	0.832 (KNN)	0.812 (KNN)	0.846 (KNN)	0.842 (KNN)
Course D (<i>n</i> = 481)	Intervention 1	0.901 (LR)	0.946 (LR)	0.992 (LR)	0.905 (LR)	0.824 (DT)	0.783 (DT)	0.824 (DT)	0.788 (DT)
Course E (<i>n</i> = 87)	Intervention 1	0.941 (SVM)	0.969 (SVM)	0.988 (SVM)	0.951 (SVM)	0.906 (KNN)	0.893 (KNN)	0.906 (KNN)	0.89 (KNN)

*KNN is (k-Nearest Neighbors), LR is (Logistic Regression), DT is (Decision Trees) and SVM is (Support Vector Machine)

***n* represents the number of historical student records used to develop the predictive models

Table 8 Comparison of control and experimental groups for at-risk students

Course	Cohort	N	Intervention 1	Intervention 1 Week	Intervention 2	Intervention 2 Week	Unique	Pass	Fail	Pass/Fail Difference	Chi-square (X^2)	P-value (Chi-Square)	P-value (Fisher's exact test)	Mean Final Mark (Standard Deviation)	P-value (t-test)
Course A	2022	57	4	6	8	10	11	9 (81.8%)	2 (18.2%)	41.8%	2.313	0.128	N/A	60.81 (20.53)	0.041
	2021	35	5	6	7	10	10	4 (40%)	6 (60%)					43.44 (14.58)	
Course B	2022	82	27	10	-	-	27	18 (66.7%)	9 (33.3%)	35.9%	6.908	0.009	N/A	58.10 (17.85)	0.000001
	2021	83	39	10	-	-	39	12 (30.8%)	27 (69.2%)					32.87 (19.94)	
Course C	2022	51	27	4	11	11	28	18 (64.3%)	10 (35.7%)	28.8%	3.799	0.051	N/A	42.37 (22.43)	0.258
	2021	51	28	4	29	11	31	11 (35.5%)	20 (64.5%)					36.00 (20.14)	
Course D	2022	71	9	6	-	-	9	1 (11.1%)	8 (88.9%)	11.1%	N/A	N/A	1.0*	14.66 (23.44)	0.939
	2021	83	9	6	-	-	9	0 (0%)	9 (100%)					15.39 (15.59)	
Course E	2022	33	4	11	-	-	4	0 (0%)	4 (100%)	0%	N/A	N/A	-*,**	19.00 (15.55)	0.840
	2021	33	6	11	-	-	2	0 (0%)	2 (100%)					15.00 (21.21)	

*Fisher's exact test was used

** No variation between pass/fail variables

The chi-square (χ^2) contingency test was used to compare the performance rates (i.e., pass and fail rates) across the two groups. The test showed that the association between the grade outcomes (pass and fail) and the treatment was not significant: χ^2 ($df=1, N=21$) = 2.313, $p=0.128$ at 5% level of significance.

The descriptive analysis of the experimental and control groups shows that the mean final mark for the course in the experimental group is 60.81 (SD=20.53) and for the control group is 43.44 (SD=14.58). The independent t-tests indicated that the difference between the means of the final marks of the experimental and control groups is statistically significant (conditions: $t(17.71)=2.20, p=0.041$). There is a 17% increase in mean final mark in the experimental group compared to the control group which is statistically significant.

Course B: Academic for course B intervened once during the semester, in Week 10 after Assignment 2. In the experimental group, 27 students were identified as at-risk, while 39 students in the control group were predicted as at-risk. The chi-square (χ^2) contingency test was used to compare the pass and fail rates across the two groups. There was strong evidence of an association between the grade outcomes and the treatment (χ^2 ($df=1, N=66$) = 6.908, p -value = 0.009 at 5% level of significance).

The experimental group had a significantly higher pass rate (66.7% in the experimental group vs. 30.8% in the control group), resulting in a 35.9% higher pass rate. The fail rates of the experimental group decreased by 35.9% compared to the control group, with 9 students (33.3% fail rate) failing the course in the treatment group, and 27 students (69.2% fail rate) failing in the control group.

The descriptive analysis of the experimental and control groups showed that the mean final mark for the course in the experimental group was 58.10 (SD=17.85) and 32.87 (SD=19.94) for the control group. The independent t-tests indicated the difference between the means of the final marks of the experimental and control groups were statistically significant (conditions: $t(59.79)=5.37, p=0.000001$). There is a 25% increase in mean final mark in the experimental group compared to the control group which is statistically significant.

Course C: In course C, the instructor intervened twice during the teaching period, in Week 4 after Practical Programming Test and Week 11 after Programming Assignment. In the experimental group, 27 students were identified as at-risk after the Practical Programming Test, while 28 students in the control group were predicted as at-risk. In the second prediction, after Programming Assignment in week 11, 11 students were identified as at-risk of failure in the experimental group, while 29 students in the control group were identified as at-risk. There were 28 unique students in the experimental group, and 31 unique students in the control group.

The chi-square (χ^2) contingency test was used to compare the pass and fail rates across the two groups. There was strong evidence of an association between the grade outcomes and the treatment at 10% level of significance (χ^2 ($df=1, N=59$) = 3.799, p -value = 0.051).

The experimental group had a significantly higher pass rate (64.3% in the experimental group vs. 35.5% in the control group), resulting in a 28.8% higher

pass rate. The fail rates of the experimental group decreased by 28.8% compared to the control group, with 10 students (35.7% fail rate) failing the course in the treatment group, and 20 students (64.5% fail rate) failing in the control group.

The descriptive analysis of the experimental and control groups showed that the mean final mark for the course in the experimental group was 42.37 (SD=22.43) and 36.00 (SD=20.14) for the control group. The independent t-tests indicated that the difference between the means of the final marks of the experimental and control groups were not statistically significant (conditions: $t(54.59) = 1.14, p = 0.258$).

Course D: In course D, instructor decided to intervene once in week 6 after Practical Test 1. After Practical Test 1 in week 6, 9 at-risk students were identified in the experimental group and 9 in the control group.

The experimental group had a higher pass rate of 11.1% compared to 0% in the control group, resulting in an 11.1% increase in the experimental group's pass rate. The fail rate in the experimental group decreased by 11.1% compared to the control group. Eight students (out of 9) in the experimental group failed the course (88.9% fail rate) while nine students (out of 9) failed in the control group (100% fail rate).

Given the small sample sizes in both groups ($n_{2021} = 9$ and $n_{2022} = 9$), Fisher's exact test was employed instead of the chi-square test to assess the association between pass and fail rates across the groups. The test results showed no statistically significant association between grade outcomes and the treatment, indicated by the p -value from the Fisher's exact test (p -value = 1.0).

Descriptive analysis of the experimental and control groups showed that the mean final marks for the course in the experimental and control groups were 14.66 (SD=23.44) and 15.39 (SD=15.59), respectively. An independent t-test showed that there was no statistically significant difference between the means of the final marks of the experimental and control groups (conditions: $t(16) = 0.078, p = 0.939$).

Course E: Instructor 4 decided to intervene once in Week 11 after Assignment 2. Given the small cohort, Instructor 4 did not use the predictive models, rather decided to intervene students who have not submitted either Assignment 1 and/or 2. Same criteria was used to identify at-risk students group in 2021 cohort (control group). Four students were identified to be at-risk in the experimental group while two students were identified as risk in the control group. Similar to Course D, this course also had small sample sizes in each group ($n_{2021} = 2$ and $n_{2022} = 4$), prompting the utilization of Fisher's exact test for comparing pass and fail rates. Fisher's exact test results for Course E cannot be calculated due to the lack of variation between control and intervention groups; as all students in both groups failed, preventing the assessment of any association or independence between the Pass/Fail variables.

The descriptive analysis of the experimental and control groups revealed mean final marks for the course to be 19.00 (SD=15.55) and 15.57 (SD=21.21) for the experimental and control groups, respectively. An independent t-test demonstrated no significant difference between the mean final marks of the experimental and control groups.

Results – Sub RQ3

To address sub RQ3, we compare metrics for the experimental group and a control group which was selected using PSM. The results of this comparison are provided in Table 9.

The chi-square (χ^2) contingency test was used to compare the performance rates (i.e., pass and fail rates) across the experimental and control groups. For all courses (Course A, B, C, D and E), the association between the grade outcomes (pass and fail) and the treatment was not statistically significant (i.e. $p > 0.05$).

The independent t-tests for courses B and E indicated that the difference between the means of the final marks of the experimental and control groups is statistically significant ($p < 0.05$). For course B, the descriptive analysis of the experimental and control groups shows that the mean final mark for the course in the experimental group is 59.9024 ($SD = 22.32289$) and for the control group is 51.9390 ($SD = 19.94582$). For course E, the descriptive analysis of the experimental and control groups shows that the mean final mark for the course in the experimental group is 68.1515 ($SD = 22.18688$) and for the control group is 57.6970 ($SD = 19.51680$). The independent t-tests for courses A, C and D indicated that the difference between the means of the final marks of the experimental and control groups is not statistically significant ($p > 0.05$).

Results – sub RQ4

We invited academics who used SPPA to obtain their views. Three academics (i.e. Instructor 2, 3 and 4) who used the SPPA in their courses participated in interviews. Interview questions focused on features of SPPA, views on using SPPA and possible limitations/improvements to SPPA. A thematic analysis (Braun & Clarke, 2006) was performed on the transcriptions using NVIVO (qualitative analysis tool) to identify and analyse themes and patterns.

A number of themes emerged from the analysis of interview data:

- (i) *Theme 1: Course design.* Instructors 2 and 3 felt that the course design feature and mapping of CLOs to ATs and TLOs was useful and helped them reflect on the course design.
- (ii) *Theme 2: Student performance prediction.* Academics found the student performance prediction model feature useful to identify at-risk and borderline students seamlessly early on and contact them. Some interventions included providing extensions to assessments (Instructor 2) and extensions with late penalty (Instructors 3).
- (iii) *Theme 3: Intervention through personalised feedback.* Academics found the personalised feedback feature useful, as students were more responsive, and was able to offer tailored assistance and interventions based on student needs.
- (iv) *Theme 4: Ease of use and usability.* Although there is an initial learning curve on SPPA's process, academics reported the system's features being intuitive to use. Instructors 2 noted that using the system for a second time was effortless

Table 9 Control vs experimental group (the entire cohort)

Course	Cohort	N	Pass	Fail	Pass/Fail Dif- ference	Chi-square (χ^2)	P-value (Chi-Square)	Mark Mean (Standard Deviation)	P-value (t-test)
Course A	2022	57	55 (96.5%)	2 (3.5%)	7%	1.210	0.271	75.7719 (14.56269)	0.281
	2021	57	51 (89.5%)	6 (10.5%)				72.2105 (20.12438)	
Course B	2022	82	64 (78.0%)	18 (22.0%)	10.9%	1.960	0.162	59.9024 (22.32289)	0.009
	2021	82	55 (67.1%)	27 (32.9%)				51.9390 (19.94582)	
Course C	2022	51	38 (74.5%)	13 (25.5%)	13.7%	1.613	0.204	56.0000 (22.48911)	0.848
	2021	51	31 (60.8%)	20 (39.2%)				55.2549 (16.21338)	
Course D	2022	71	56 (78.9%)	15 (21.1%)	1.4%	0.0	1.0	60.2254 (24.42317)	0.362
	2021	71	55 (77.5%)	16 (22.5%)				57.0423 (16.19386)	
Course E	2022	33	28 (84.8%)	5 (15.2%)	3%	0.0	1.0	68.1515 (22.18688)	0.046
	2021	33	27 (81.8%)	6 (18.2%)				57.6970 (19.51680)	

because previous data (CA Mapping model data and predictive models) were already available in the system. However, academics suggested integration SPPA (as a feature) with the LMS to minimise the use of separate systems. The feedback suggests this feature would be important for SPPA's adoption.

- (v) *Theme 5: Academic dashboards and course evaluation dashboards.* Academics found dashboards helpful for tracking student progress and benchmarking against past performances. Instructor 2 emphasised their necessity for monitoring course progress and making the comparisons against previous cohorts.
- (vi) *Theme 6: Effects of interventions.* Academics felt that they are able to identify student issues and provide targeted interventions by using SPPA. For instance, Instructors 4's interventions helped some students pass the course and resolved technical assessment submission issues for certain students.
- (vii) *Theme 7: Overall helpful and useful.* Academics found the system helpful in refining course design and boosting student performance via personalised feedback and predictive capabilities. They appreciated the opportunity to closely monitor students' progress and make data-driven decisions to intervene with and support struggling students.

Overall, academics found SPPA's process and features useful and intuitive to use. There were a couple of areas identified that needed to further improvements. The need to login to a different system other than the LMS was considered a major drawback and also a reason for not using the student dashboard feature which requires all students to login to a different system. Also, the fact that the data being hosted on a non-institutional system was raised as a concern for data security and privacy.

Results – sub RQ5

Instructor 4 used SPPA and provided interventions in Course F in 2021 and in 2022. Course F was an online course with a small enrolment of students. Since Course F had a small number of enrolments, the academic did not use predictive models, rather send intervention emails based on students' progress in continuous assessment tasks. Table 10 outlines the interventions undertaken by the academic. Note that given that Course F was offered online, the academic was able to view the progress each student made on their Module Tasks – which was based on activities conducted using the LMS—even before the Module Tasks were due. In 2021, the academic decided to intervene in week 7 after Practical Test 1 and Module 1 Weekly Tasks are due. In 2022, the academic decided to intervene four times based on their progress on assessment tasks at different times (i.e. Week 3, 4, 5 and 8) of the academic term.

The results of using SPPA in 2021 and 2022 cohorts in course F for the intervened students is provided in Table 11 and for the entire cohort is provided in Table 12. Note that the students who dropped are not included in the results. Due to the small size of intervened students, in both groups ($n_{2021}=9$ and $n_{2022}=2$), Fisher's exact test was employed instead of the chi-square test to assess the association between pass and fail rates across the groups. The test results showed no statistically significant association between grade outcomes and the treatment, indicated by the p -value from the Fisher's exact test (p -value = 0.467). The independent t-tests for the

Table 10 Interventions in Course F

Course	Cohort	Intervention	Week	Description
Course F	2021	Intervention 1	Week 7	Students who did not submit Practical Test 1 and who did not submit or performed poorly in Module 1 Weekly Tasks
	2022	Intervention 1	Week 3	Not shown any progress in Module 1 Weekly Tasks
		Intervention 2	Week 4	Before Module 1 due date. Students who have had no progress in Module 1
		Intervention 3	Week 5	Students who have not submitted Module 1 Weekly Tasks
		Intervention 4	Week 8	Students who have not submitted assessment tasks up to Module 2 Weekly Tasks

Table 11 SPPA in two iterations: intervened cohorts (Course F)

Course	Cohort	N	Intervention 1	Intervention 2	Intervention 3	Intervention 4	Unique	Pass	Fail	Pass/Fail Difference	P-value (Fisher's exact test)	Mark Mean (Standard Deviation)	P-value (t-test)
Course F	2022	27	14	6	2	2	14	8 (57.1%)	6 (42.9%)	57.14%	0.467	47.1429 (28.19945)	0.000017
	2021	23	2	-	-	-	2	0 (0%)	2 (100%)			1.0000 (1.41421)	

Table 12 SPPA in two iterations: entire cohort (Course F)

Course	Cohort	N	Pass	Fail	Pass/Fail Difference	Chi-square (χ^2)	P-value (Chi-Square)	Overall Mark Mean (Standard Deviation)	P-value (t-test)
Course F	2022	27	17 (65.4%)	9 (34.6%)	(12.9%)	0.461	0.497	55.9231 (25.41956)	0.448
	2021	23	18 (78.3%)	5 (21.7%)				61.8261 (28.26442)	

two cohorts indicate that the difference between the means of the final marks of the 2021 and 2022 cohorts are statistically significant (i.e., $p < 0.05$).

The results comparing the entire cohort of students for Course F in 2021 and 2022 are provided in Table 12. The chi-square (χ^2) contingency test is used to compare the performance rates (i.e., pass and fail rates) across the two cohorts. The association between the grade outcomes (pass and fail) for the two cohorts are not statistically significant (i.e., $p > 0.05$). The independent t-tests for the two cohorts indicate that the difference between the means of the final marks of the 2021 and 2022 cohorts are not statistically significant (i.e., $p > 0.05$).

Limitations, Discussion and Future Research Directions

It is important to note that there are a number of limitations of this study. A relatively small number of courses and diversity of courses were considered in the study (i.e. 6 courses with 5 in blended mode and 1 online). The size of the cohorts in courses where SPPA ranged from 23–82. It is expected that LAIs would most likely benefit in courses with large number of students where it is impractical for instructors to personally keep track of students and their progress. It would be desired to apply SPPA to larger cohorts (e.g. cohorts ranging from 100s – 100,000 students such as in MOOC courses) and evaluate their impact. The courses are of technical nature – physics, computing. It would be good to consider courses that are in a variety of disciplines (e.g. arts, social sciences, project-based courses, and others). Thus, it is important to consider these limitations when generalizing conclusions from the results of this study.

The results addressing sub RQ1 validate that continuous historical assessments data of a course can be used to develop high-performing course-specific student performance predictive models. This is in line with our previous work (Alalawi et al., 2021b and Alalawi et al., 2024). SPPA's use of historical continuous assessment data of courses to develop course-specific predictive models academics to develop predictive models without the need to access data from various educational IT systems. The limitation of this approach is that predictions can be performed only after the students have submitted a continuous assessment in the course. Another limitation of this approach is that as assessments changes between course iterations as courses evolve (e.g. adding a new assessment), SPPA may not be able to create predictive models for the new assessments in its first course iteration (e.g. predictive models may not be available for new assessments until historical records are available). Future work can consider incorporating research where predictive models with evolving features can be created (e.g. Hou et al., 2022) to keep up with changing course assessment structures.

With SPPA's flexible access via a web browser and positive acceptance by academics to SPPA's approach and its features to pilot LAIs (see sub RQ4 results), SPPA has the potential to influence the uptake and adoption of LAIs (addressing RQ1). However, from the qualitative results from academics (results of sub RQ4), it was clear that academics preferred to have SPPA integrated to LMS as a feature as this would allow academics and students to seamlessly access SPPA and its features.

This approach would also eliminate the need to upload continuous assessment data manually to create predictive models as the LMS would already have access to continuous assessment data of previous course iterations. Also, LMS integration can allow SPPA to access other related data which can be used to create predictive models (e.g. student engagement data in LMS, etc.). This approach also can address the limitation where predictions can be undertaken even before the first continuous assessment is due in the course (e.g. Milliron et al., 2014). Future work can consider developing SPPA as a LTI² module that integrates into LMS and also consider integrating related data sets to create predictive models. Such work in the future has the potential to have a significant positive impact and influence in the uptake and adoption of SPPA.

In the results for sub RQ2, we observe that statistically significant improvements in final grades in Course A (17% increase) and Course B (25% increase) as well as improved pass rates in Course B (35% increase) and C (28% increase). However, there are no observable (i.e. statistically significant) changes in pass rates for course A, D and E as well as observable (statistically significant) changes in final grades for C, D and E. In the results for sub RQ3, Course B and E had a statistically significant improvements in final grades (8% increase in Course B and 10% increase in Course E) in the experimental group. Overall, these results demonstrates that SPPA can facilitate effective interventions (addressing sub RQ2 and sub RQ3) but observable improvements (i.e. statistically significant) are not guaranteed in all cases. Also, the feedback from academics further validates that SPPA's features (such as mapping CLOs, TLAs and ATs during course design, etc.) are deemed useful and relevant to guide effective interventions. What is also evident is that using SPPA itself only do not guarantee observable (i.e. statistically significant) improvements. Other factors are important and plays a role (e.g. what actions are taken during intervention, when, how). In this study, the academics mainly sent personalised email interventions to alert students who are at risk, request for an update/status, encourage them to seek help and plan for future assessments. In some courses, when students responded, academics provided opportunity to submit missed assessments (e.g. Course B), extend due date with late penalty (e.g. Course E), etc. These actions, in addition to personalised email interventions, had improved outcomes for students who took advantage of these opportunities. Of course, some students did not respond to intervention emails. This may be due to a number of reasons such as these students not checking emails, may have already given up attempting to pass the course by the time of interventions, and others. In such situations, interventions may not help and may be too late.

The results of sub RQ5, demonstrated that even if SPPA is used in multiple iterations of the course it cannot guarantee improvements. In sub RQ5 results, 2021 cohort was subject to a single intervention, while in 2022 was subject to four iterations. There were no observable (statistically significant) changes between the entire cohort (pass/fail rates and mean final marks) while for the intervention cohort, there was an observable (statistically significant) improvement in mean final marks in

² <https://www.imsglobal.org/spec/lti/v1p3>

2021 when compared to 2022 cohort. These results outline the challenges academics face when deciding what types of interventions, when and under what conditions is ideal.

The above results do re-enforce discussions in literature. Interventions are context-sensitive and is a challenge to identify the “optimal” interventions (Rienties et al., 2017). As Milliron et al. (2014) asserts, there are no “silver bullet” to interventions and it is an iterative process. Although, SPPA provides information to assist in interventions (e.g. gaps in student knowledge, personalised revision plans) and aims to guide interventions using sound pedagogical approaches (e.g. Hattie and Timperley’s (2007) model for effective feedback), the interventions themselves were left for the discretion of academics. Future research in LAIs can consider further assistance for academics to help guide in identifying effective ways to guide interventions.

We also observe that in Borrella et al. (2022), interventions A and B which provided motivational email communication for at-risk students prior to an assessment or preparation and study guidelines before an exam for at-risk students of dropping out students did not result in improved observable outcomes. However, in interventions C and D in Borrella et al. (2022), where “problem” areas in a particular course iteration were identified (e.g. perceived difficulty of assessments in Intervention C, difficulty of course content in Intervention D) and then applying an intervention strategy focused on addressing the problem/issue identified using a didactic scaffolding learning approach (i.e. sound pedagogical approach) improved outcomes. This approach has the potential to lead to a continuous improvement model if “problem” areas are identified in a course iteration and appropriate intervention strategies can be designed to address these “problem” areas in the next course iteration.

SPPA’s approach which considers the entire course life cycle and multiple course iterations lends itself to this continuous improvement model (i.e. addressing RQ3). In the course evaluation phase, “problem” areas to address can be identified. Tools such as SPPA’s course evaluation dashboards and supplementary course evaluation data (e.g. feedback from students, academics, etc.) can be helpful to identify “problem” areas. Next, once problems/issues are identified, interventions can be designed in the course design phase of the next course iteration taking course/program-related factors (institutional factors) (Lee & Choi, 2011) into consideration and using sound pedagogical principles to guide development of effective interventions addressing the identified “problem” areas. These interventions are applied in the next course iteration and evaluated in the course evaluation phase for the impacts. This approach can be repeated to develop a continuous improvement model (see Fig. 10). Of course, the challenge is to identify relevant “problems/issues” in a course iteration and then decide on the most appropriate intervention strategy to address them (i.e. no “silver bullet”). Although dashboards (such as course evaluation dashboards in SPPA can help along with supplementary course evaluation data – student feedback etc.), this information needs to be critically analysed to identify “problem” areas or “potential” areas for improvement. Next, appropriate effective interventions needs to be designed. The use of sound pedagogical principles or learning theories may be able to guide the design of such interventions. We assume that such an approach will require time, effort, collaboration and resources from relevant stakeholders including support from organisation’s culture for such endeavours. Future research work

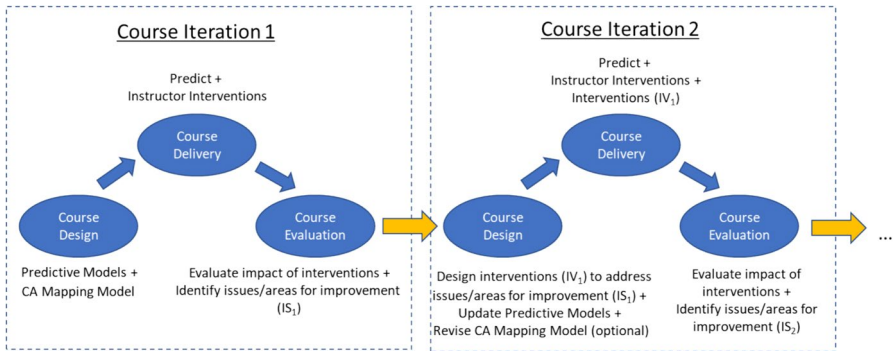


Fig. 10 A proposed model for continuous monitoring, evaluating and improving in LAIs

needs to undertaken to further investigate, validate, fine-tune and improve such models for a continuous improvement model in LAIs. We also note that recent advances in Artificial Intelligence (AI) may be considered in future to assist in identifying and selecting the most appropriate intervention strategies for students/cohorts (e.g. analysing historical interventions strategies and their success across students/cohorts and suggesting which strategies have better chance for success for particular students in the current cohort).

Conclusion

Predicting student performance has gained much attention in literature where diverse student related data sets are collected and analysed typically using data mining and ML techniques to predict the students' performance levels. If at-risk students are known early on, stakeholders (academics, administrators and students) can undertake appropriate interventions to improve educational outcomes. In literature, a lot of work on student performance prediction models exists but fewer studies have considered actions that can be taken based on these prediction results and evaluating their impact. LAI studies address this gap. In LAIs, predictive models for student performance are developed (typically integrating diverse relevant data set and using EDM/ML techniques) and this information is disseminated to relevant stakeholders for targeted interventions to improve educational outcomes. Typically, LAI infrastructure includes not only predictive models but also tools for information dissemination (such as LADs) and personalised communication and templates for communication to assist in interventions. In literature, LAI studies have shown significant potential for positive outcomes (improved pass rates, improved grades, reduced dropouts etc.).

Although LAI studies have been shown to improve outcomes as early as in 2012, their uptake of LAIs has been slow and is still in its infancy. This paper identifies the main impediments to the uptake of LAIs. Lack of access to LAI infrastructure is one the main reasons why academics cannot pilot LAIs in their courses. LAI infrastructure requires developing customized predictive models for an institution taking related data from various IT systems. This requires extensive

institution-centric efforts and investments. Another challenge is providing effective interventions. In LAIs, the interventions themselves are under the discretion of academics and other stakeholders who are aware of the learning context and in control of decision-making. Thus, the effectiveness of interventions is dependent on the skills, experience and knowledge of the academics and other stakeholders with no guarantee of success. Also, there is little research or evidence on continuous improvement after LAIs have been applied in courses. Given these reasons, institutions may not be incentivised to prioritise investments in LAIs even with its potential for significant impacts.

This paper presents a framework, termed SPPA, which allows academics to pilot LAIs in their courses accessing LAI infrastructure without the need to integrate diverse data sets from existing IT systems in educational institutions, avoiding the need for large scale investments. SPPA promotes a self-service model to pilot LAIs by academics in their courses accessing LAI infrastructure and related features conveniently via a web browser. Academics develop course-specific student performance predictive models using just continuous assessments data of the course from previous iterations. SPPA uses different ML algorithms to create predictive models and chooses the best performing models based on a number of metrics. SPPA provides a number of tools such as dashboards to disseminate relevant information for monitoring and evaluating courses and students' progress as well as tools to integrate personalised email interventions. In SPPA, the specific interventions are under the discretion of the academic, however, SPPA guides and assists academics to develop effective interventions. SPPA incorporates sound pedagogy principles in course design and providing feedback, along with information such as students' knowledge gaps and personalised revisions plans to assist and guide academics to provide effective, personalised and targeted feedback and interventions. SPPA considers the entire lifecycle of courses from course design, delivery and evaluation iteratively that can be extended to provide a model for continuous improvement in LAIs. SPPA has the potential to fast-track the adoption and uptake of LAIs addressing one of the main obstacles (i.e. access to LAI infrastructure by academics without the need for large-scale institutional investments). Also, SPPA's code provided as an open-source project for further development and use.

SPPA was evaluated in a tertiary educational institution in six courses. Both qualitative and quantitative data was collected and analysed. The predictive models generated using continuous assessment data was shown to be effective using a number of metrics. Academics used SPPA to pilot LAIs and evaluated its ability to facilitate effective interventions, impact on overall cohort and also evaluate SPPA's usability and features. Quasi-experimental approach was designed to evaluate the impact of interventions and the entire cohort. Statistical tests were conducted on experimental group and control groups on pass rates, fail rates and mean final marks to evaluate any observable (i.e. statistically significant) impact. Qualitative data was collected and analysed from academics on their opinion of using SPPA to provide LAIs. These results demonstrated that SPPA's ability to provide a self-service model for academics to pilot LAIs and provide effective interventions. Overall, academics who used SPPA had a positive outlook on SPPA and its features and found it helpful to improve outcomes.

There were many areas for future investigation that were identified. Feedback from academics clearly outlined that seamless access to SPPA via LMS would be highly desirable feature. This approach would also allow SPPA to access data sets from LMS without manual intervention and also allow improving predictive models using related existing data sets from the LMS. The evaluation results demonstrated that observable (i.e. statistically significant) improvement by LAIs is not always guaranteed. That is, finding optimal interventions in LAIs still remains a challenge. Further research on how to develop effective interventions – what works, when and under what conditions is needed. A model for continuous improvement in LAIs was outlined whereby areas for improvement are identified (in a course iteration) and then interventions designed with sound pedagogical approaches to address these identified areas in the next course iteration, that can lead to a higher potential for success and also support a continuous improvement process model in LAIs. Future research work is needed to validate these models.

Acknowledgements We would like to thank Ms H.C. Fernando for her input in improving the statistical analysis.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability The authors do not have permission to share data.

Declarations

Conflict of Interest The authors have no competing interests to report for this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alalawi, K., Athauda, R., & Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. *Engineering Reports*, 5(12), e12699. <https://doi.org/10.1002/eng2.12699>
- Alalawi, Athauda, R., & Chiong. (2021a). *An Innovative Framework to Improve Course and Student Outcomes*. Paper presented at the 2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA).
- Alalawi, Chiong, & Athauda. (2021b). *Early Detection of Under-Performing Students Using Machine Learning Algorithms*. Paper presented at the CITISIA 2021 - IEEE Conference on Innovative Technologies in Intelligent System and Industrial Application, Proceedings.
- Alalawi, K., Athauda, R., Chiong, R., & Renner, I. (2024). Evaluating the student performance prediction and action framework through a learning analytics intervention study. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12923-5>
- Anderson, L. W., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*: Longman.

- Arnold, K. E., & Pistilli, M. D. (2012). *Course signals at Purdue: using learning analytics to increase student success*. Paper presented at the Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, British Columbia, Canada. <https://doi.org/10.1145/2330610.2330666>
- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Baneres, D., Rodríguez-Gonzalez, M. E., & Serra, M. (2019). An Early Feedback Prediction System for Learners At-Risk Within a First-Year Higher Education Course. *IEEE Transactions on Learning Technologies*, 12(2), 249–263. <https://doi.org/10.1109/TLT.2019.2912167>
- Baneres, D., Rodríguez, M. E., Guerrero-Roldán, A. E., & Karadeniz, A. (2020). An early warning system to detect at-risk students in online higher education. *Applied Sciences*, 10(13), 4427.
- Baneres, D., Rodríguez-González, M. E., Guerrero-Roldán, A.-E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, 20(1), 1–25.
- Begum, S., & Padmanavar, S. S. (2022). Genetically Optimized Ensemble Classifiers for Multiclass Student Performance Prediction. *International Journal of Intelligent Engineering and Systems*, 15(2), 316–328. <https://doi.org/10.22266/ijies2022.0430.29>
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. *Office of Educational Technology, US Department of Education*.
- Biggs. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education*, 1(1), 5–22.
- Biggs, J. B., & Collis, K. F. (2014). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2022). Taking action to reduce dropout in MOOCs: Tested interventions. *Computers & Education*, 179, 104412. <https://doi.org/10.1016/j.compedu.2021.104412>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Burgos, C., Campanario, M. L., de la Pena, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering*, 66, 541–556.
- Cambruzzi, W. L., Rigo, S. J., & Barbosa, J. L. (2015). Dropout prediction and reduction in distance education courses with the learning analytics multitrail approach. *Journal of Universal Computer Science*, 21(1), 23–47.
- Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26(19), 3661–3675. <https://doi.org/10.1002/sim.2832>
- Choi, S. P., Lam, S. S., Li, K. C., & Wong, B. T. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society*, 21(2), 273–290.
- Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6), 683–695. <https://doi.org/10.1080/13562517.2013.827653>
- Corrigan, O., Smeaton, A. F., Glynn, M., & Smyth, S. (2015). *Using educational analytics to improve test performance*. Paper presented at the Design for Teaching and Learning in a Networked World.
- Dawson, S., Jovanović, J., Gašević, D., & Pardo, A. (2017). *From prediction to impact: Evaluation of a learning analytics retention program*. Paper presented at the Proceedings of the seventh international learning analytics & knowledge conference.
- Dodge, B., Whitmer, J., & Frazee, J. P. (2015). *Improving undergraduate student achievement in large blended courses through data-driven interventions*. Paper presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, Poughkeepsie, New York. <https://doi.org/10.1145/2723576.2723657>
- Earl, W. R. (1988). Intrusive Advising of Freshmen in Academic Difficulty. *NACADA Journal*, 8(2), 27–33.

- Enaro, A. O., & Chakraborty, S. (2020). Feature selection algorithms for predicting students academic performance using data mining techniques. *International Journal of Scientific and Technology Research*, 9(4), 3622–3626.
- Espinoza, P., & Genna, G. M. (2021). Hi, I Want to Talk to You About Your Progress: A Large Course Intervention for At-Risk College Students. *Journal of College Student Retention: Research, Theory & Practice*, 23(1), 2–27. <https://doi.org/10.1177/1521025118790054>
- Fang, B., & Shewmaker, J. (2016). *The Case for Small Data in Higher Education*. Paper presented at the Australasian Conference on Interactive Entertainment.
- Fasihuddin, H., Skinner, G., & Athauda, R. (2017). Towards adaptive open learning environments: Evaluating the precision of identifying learning styles by tracking learners' behaviours. *Education and Information Technologies*, 22(3), 807–825. <https://doi.org/10.1007/s10639-015-9458-5>
- Figuroa-Cañas, J., & Sancho-Vinuesa, T. (2021). Changing the recent past to reduce ongoing dropout: an early learning analytics intervention for an online statistics course. *Open Learning: The Journal of Open, Distance and e-Learning*, 39(3), 194–211. <https://doi.org/10.1080/02680513.2021.1971963>
- Foster, C., & Francis, P. (2020). A systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes. *Assessment & Evaluation in Higher Education*, 45(6), 822–841. <https://doi.org/10.1080/02602938.2019.1696945>
- Gajwani, J., & Chakraborty, P. (2021). *Students' Performance Prediction Using Feature Selection and Supervised Machine Learning Algorithms*. Paper presented at the 3rd International Conference on Innovative Computing and Communication, ICICC 2020.
- Garg, S., Aleem, A., & Gore, M. M. (2021) Employing Deep Neural Network for Early Prediction of Students' Performance. In: *Vol. 185 LNNS. Lecture Notes in Networks and Systems* (pp. 497–507).
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68–84.
- Guerrero-Roldán, A.-E., Rodríguez-González, M. E., Baneres, D., Elasmri-Ejjaberi, A., & Cortadas, P. (2021). Experiences in the use of an adaptive intelligent system to enhance online learners' performance: A case study in Economics and Business courses. *International Journal of Educational Technology in Higher Education*, 18, 1–27.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hou, B. J., Zhang, L., & Zhou, Z. H. (2022). Prediction With Unpredictable Feature Evolution. *IEEE Transactions on Neural Networks and Learning Systems*, 33(10), 5706–5715. <https://doi.org/10.1109/TNNLS.2021.3071311>
- Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, 1(1), 6–47. <https://doi.org/10.18608/jla.2014.11.3>
- Keller, J. M. (1987). The systematic process of motivational design. *Performance and Instruction*, 26, 1–8.
- Kolb, D. (1984). *Experiential Learning: Experience As The Source Of Learning And Development* (Vol. 1).
- Kovacic, Z. (2010). *Early prediction of student success: Mining students' enrolment data*. Paper presented at the Informing Science + Information Technology Education Joint Conference, Cassino, Italy.
- Krajcik, J. S., & Blumenfeld, P. C. (2006). *Project-based learning*. Cambridge University Press.
- LarrabeeSønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594–2618. <https://doi.org/10.1111/bjjet.12720>
- Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, 59(5), 593–618. <https://doi.org/10.1007/s11423-010-9177-y>
- Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, 47, 90–97. <https://doi.org/10.1016/j.chb.2014.07.013>
- Lu, O. H., Huang, J. C., Huang, A. Y., & Yang, S. J. (2018). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. In *Learning Analytics* (pp. 78–92): Routledge.

- Lune, H., & Berg, B. L. (2021). *Qualitative research methods for the social sciences* (9th edition ed.): Pearson.
- Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining techniques applied in educational environments: Literature review. *Digital Education Review*, 33, 235–266.
- Maybin, J., Mercer, N., & Stierer, B. (1992). Scaffolding learning in the classroom. *Thinking Voices: The Work of the National Oracy Project*, 186, 195.
- Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and Action Analytics: Three Case Studies to Consider. *Research & Practice in Assessment*, 9, 70–89.
- Mohamad, S. K., & Tasir, Z. (2013). Educational Data Mining: A Review. *Procedia - Social and Behavioral Sciences*, 97(2013), 320–324. <https://doi.org/10.1016/j.sbspro.2013.10.240>
- Mubarak, A. A., Cao, H., & Zhang, W. (2022). Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, 30(8), 1414–1433. <https://doi.org/10.1080/10494820.2020.1727529>
- Pardo, A., Bartimote, K., Shum, S. B., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martínez-Maldonado, R., Mirriahi, N., & Moskal, A. C. (2018). OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Journal of Learning Analytics*, 5(3), 235–249. <https://doi.org/10.18608/jla.2018.53.15>
- Rahal, A., & Zainuba, M. (2016). Improving students' performance in quantitative courses: The case of academic motivation and predictive analytics. *The International Journal of Management Education*, 14(1), 8–17. <https://doi.org/10.1016/j.ijme.2015.11.003>
- Ramaswami, G., Susnjak, T., & Mathrani, A. (2022). On Developing Generic Models for Predicting Student Outcomes in Educational Data Mining. *Big Data and Cognitive Computing*, 6(1), 235. <https://doi.org/10.3390/bdcc6010006>
- Rao, J. N. K., & Scott, A. J. (1984). On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. *The Annals of Statistics*, 12(1), 46–60. <http://www.jstor.org/stable/2241033>.
- Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a Learning Analytics Intervention and Evaluation Framework: What Works? In B. Kei Daniel (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (pp. 147–166). Cham: Springer International Publishing.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30–41. <https://doi.org/10.1109/TLT.2016.2599522>
- Slater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. *London: Jisc. Accessed February*, 8 (2017) 176.
- Song, Z., Sung, S.-H., Park, D.-M., & Park, B.-K. (2023). All-Year Dropout Prediction Modeling and Analysis for University Students. *Applied Sciences*, 13(2), 1143. <https://www.mdpi.com/2076-3417/13/2/1143>.
- Varney, J. (2007). Intrusive Advising. *Academic Advising Today*, 30(3), 11.
- Varney, J. (2012). Proactive (intrusive) advising. *Academic Advising Today*, 35(3), 1–3.
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*: Harvard university press.
- Wang, X., Zhang, L., & He, T. (2022). Learning performance prediction-based personalized feedback in online learning via machine learning. *Sustainability*, 14(13), 7654. <https://www.mdpi.com/2071-1050/14/13/7654>.
- Wise, A. F. (2014). *Designing pedagogical interventions to support student use of learning analytics*. Paper presented at the LAK '14: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, Indianapolis, IN, USA.
- Wong, B.T.-M., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, 7(1), 7–28. <https://doi.org/10.1007/s40692-019-00143-7>
- Wong, B. T. M. (2017). Learning analytics in higher education: An analysis of case studies. *Asian Association of Open Universities Journal*, 12(1), 21–40. <https://doi.org/10.1108/AAOUJ-01-2017-0009>

Authors and Affiliations

Khalid Alalawi^{1,2} · Rukshan Athauda¹  · Raymond Chiong^{1,3}

✉ Rukshan Athauda
Rukshan.Athauda@newcastle.edu.au

Khalid Alalawi
khalid.alalawi@uon.edu.au; kalwai@taibahu.edu.sa

Raymond Chiong
Raymond.Chiong@newcastle.edu.au

¹ School of Information and Physical Sciences, The University of Newcastle, Callaghan, Australia

² College of Computer Science and Engineering (CCSE), Taibah University, Medina, Saudi Arabia

³ School of Science and Technology, University of New England, Parramatta, Australia