Research article

Using learning analytics in higher education: Assessing students' learning experience in an actuarial science course

Lay Guat Chan* and Qian Yun Ng

School of Mathematical Sciences, Sunway University, 47500 Bandar Sunway, Selangor, Malaysia

* **Correspondence:** Email: layguatc@sunway.edu.my; Tel: +6 03-7491 8622.

Academic Editor: Siow Li Lai

Abstract: It is becoming increasingly evident that educators need to prioritize the welfare of their students, particularly those who are underperforming academically, also known as "students at risk". By analyzing learning behaviors, including attendance records, past academic results, and online interactions, we can identify students at risk and provide them with timely support. Therefore, this study aimed to develop a prediction model for identifying students at risk in an actuarial science course and suggest an intervention strategy. Our study was comprised of five components of learning analytics: data collection, reporting, prediction, intervention, and reassessment. Prior to applying a prediction model, correlation analysis was utilized to identify variables impacting students' academic performance. Three variables, including CGPA, pre-requisite subject marks, and assessment marks were considered due to their rather strong correlation with the final marks of the course. Then, quadratic discriminant analysis (QDA) was applied to predict students classified as "at risk" and "not as risk". Out of 69 students from the course, 15 students identified as "at risk" and 40 students participated in the Peer Assisted Learning Program (PALP) as an intervention strategy to reduce the course's failure rate. We cannot conclude whether PALP was an effective intervention strategy for students at risk because a majority of them failed to attend. However, we observed that those who attended PALP had a higher likelihood of passing the course. Our prediction model had high rates of accuracy, precision, sensitivity, and specificity, which were 91%, 98%, 91%, and 91%, respectively. Therefore, QDA could be considered a robust model for predicting students at risk. We have outlined some limitations and future studies at the end of our study.

Keywords: learning analytics, correlation analysis, quadratic discriminant analysis, intervention, student at risk

1. Introduction

The role of teachers in addressing student welfare, particularly in responding to underperforming students in academics (known as "students at risk"), is increasingly crucial [1]. According to Alyahyan and Düştegör [2], there are various factors that can impact a student's academic performance, such as their prior academic achievements, demographics, and online learning behaviors. If these factors can be identified earlier, educators can use the data to identify students at risk at an earlier stage.

The advancement of information and communication technology has changed the way students acquire knowledge, particularly for generations surrounded by technology from a young age. The increased use of learning management systems (LMS) has generated substantial educational data, leading to the emergence of learning analytics [3]. As a result, learning analytics is increasingly being utilized in identifying students at risk and providing them with timely assistance [4]. Instructors can monitor student behavior in real-time and intervene promptly to support students who are at risk, which ultimately helps to reduce failure rates [5].

Learning analytics refers to the process of identifying a problem, applying statistical models using data, and predicting future trends to obtain actionable insights [6]. The Signals project conducted at Purdue University is one of the most well-known learning analytics initiatives. Under the project, students are given signals in the form of red, yellow, or green on their Blackboard site. Lecturers can monitor students who receive yellow and red signals at an early stage and provide necessary interventions to help them [7]. Previous studies have demonstrated that it is possible to predict the academic performance of students in mathematics [8]. However, it is crucial to develop and validate prediction models for different courses [9], such as actuarial science, which involves different mathematics subjects from other courses.

A complete learning analytics process involves five steps: data collection, data reporting, prediction, intervention, and reassessment [10]. However, most research works have primarily focused on the first three steps and paid less attention to the last two steps [4]. Although the Signals project did provide support for students at risk, it did not specify the learning support for mathematics subjects. As pointed out in the study [8], most learning analytics relevant studies have relied on quantitative methods and emphasized that future studies can explore the use of mixed methods to support learning analytics applications in mathematics. Therefore, there is a need to bridge this gap for a comprehensive learning analytics process.

This study aimed to develop a prediction model for locating students at risk in an actuarial science course and suggest intervention strategies for them. A quantitative method was applied on the predictive model, while a qualitative method was used to gather insight on the intervention strategy. We concluded the study by assessing the effectiveness of both the predictive model and intervention strategy.

2. Literature review

2.1. Predictive model

Different classification techniques are necessary to classify students as "at risk" or "not at risk", such as discriminant analysis, logistic regression, and neural networks [11]. Statistical modelling has a high prediction accuracy of 84% [12] and possibly even 86% [13] for students at risk. Therefore, it is becoming more and more popular for classifying students at risk and predicting academic performance. One popular technique for classification and dimensionality reduction that allows for

non-linear data separation is quadratic discriminant analysis (QDA) [14].

QDA is a supervised learning technique that models the probability distribution of each class using a quadratic function [15]. It is assumed that each class's predictor variables have a multivariate normal distribution. QDA uses the predictor variables of a class to estimate the likelihood that an observation will belong to that class. Using training data, the quadratic function's means and covariance matrices are estimated for each class [15]. Using the Bayes theorem to compute the posterior probability of each class's observation, a new observation is classified using QDA by being assigned to the class with the highest probability [16].

QDA can identify more complex patterns in the data than linear methods. Because of this, it can be used in situations where predictor variables and classes have a non-linear relationship, allowing for more flexible decision boundaries [16]. Thus, QDA provides the benefit of more flexibility with respect to the features of the covariance matrix for various classes and fewer restrictive assumptions [14].

2.2. Learning behavior affecting students' academic performance

An overview of the learning behaviors influencing academic performance is given in Table 1. Lakkaraju et al. [17] found that past academic achievement, as determined by the cumulative grade point average (CGPA), is a predictor of academic performance. On the other hand, Choi et al. [4] found a relationship between exam scores in pre-requisite courses and overall academic success. Mueen et al. [13] and Yang et al. [18] stressed the significance of consistent exercise grades and homework performance, and highlighted their impact on academic achievement. The importance of Blackboard clicks in measuring engagement was highlighted by Shah and Barkas' [19] investigation of student interactions within the Blackboard learning management system.

Furthermore, Yang et al. [18] highlighted a relationship between the number of video views and academic achievement, while Mubarak et al. [12] investigated the possible impact of the total amount of time spent watching videos. Besides, six studies [4,13,17,19–21] supported the significance of attendance as a factor impacting academic performance. Finally, studies by Choi et al. and Yang et al. [4,18] show a connection between assessment performance and academic success. To summarize, the results in Table 1 show that a number of variables, including past academic performance, various forms of engagement, attendance, and assessment performance, have been studied and found to be associated with academic success.

Table 1. Literatures that discussed variables applied in prediction of students at risk.

Variables	Sources
Past academic performance (CGPA)	Lakkaraju et al. (2015)
Exam marks in the pre-requisite course	Choi et al. (2018)
Homework performance, Exercise grades	Mueen et al. (2016), Yang et al. (2018)
Blackboard clicks	Shah and Barkas (2018)
Video views	Yang et al. (2018)
Total minutes spent in videos	Mubarak et al. (2020)
Attendance	Choi et al. (2018), Mueen et al. (2016), Steward et al. (2011), Shah and Barkas (2018), Nepal and Rogerson (2020), Lakkaraju et al. (2015)
Assessment performance	Choi et al. (2018), Yang et al. (2018)

2.3. Intervention strategy

It is the duty of educational institutions to provide students at risk with intervention activities in order to lower dropout rates. For higher education, the Peer Assisted Learning Program (PALP) is regarded as a significant intervention strategy [22]. In addition to helping students transition to university life and develop better study habits, PALP has proven to be helpful by offering a secure and friendly environment for discussion with mentors [22]. Besides, according to Cheng and Walters [23], attending PALS in mathematics raises the chance that students will pass the course and complete the program.

## 3. Methodology

3.1. Participants

This study targeted full-time actuarial science undergraduate students from a private university in Selangor. A convenience sampling was applied to select the samples, where the students were selected because of the convenient accessibility of the study. Two datasets were collected for this research. The first dataset was utilized for developing the prediction model, while the second dataset was used to predict students who were at risk. Both datasets were collected from different groups of students in two semesters. The first dataset consisted of a total of 61 students, while the second dataset had a total of 69 students. All of them were enrolled in a Year 2 actuarial science course. The demographic information of the students is presented in Table 2.

Table 2. The demographic of the sample of students.

	Gender	n	Percentage
First dataset	Male	35	57.4
	Female	26	42.6
	Total	61	100.0
	Gender	n	Percentage
Second dataset	Male	47	68.1
	Female	22	31.9
	Total	69	100.0

3.2. Research procedures

In this study, we have conducted five steps of learning analytics as presented in Figure 1. The details for each step will be elaborated further in the following sections.
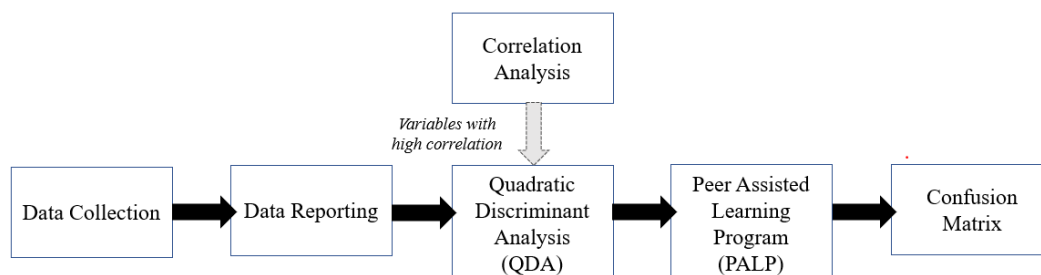
Figure 1. Research procedures of learning analytics.

3.2.1. Data collection and data reporting

The data presented in Table 3 was collected to assess the impact of different variables on students' academic performance. The primary data collected were Blackboard data, attendance, CGPA, pre-requisite subject marks, final marks of the Year 2 course, and gender. Five types of Blackboard data were collected: Blackboard clicks, videos views, total minutes spent in videos, homework marks, and assessment marks. Blackboard is one of the teaching and learning tools which has been widely used in that private university and students are required to use it from their first year of study. Attendance was collected through the university attendance system. Students' marks in pre-requisite subjects, their current CGPA, gender, and final marks were also collected from the Enterprise Manager Database Express (EMX), i.e., a system to key in and to view students' data upon their registration to join the university.

As stated in section 3.1, two datasets were collected for this research. Despite having the same variables for both datasets, their descriptive statistics differed. For this paper, we will only be discussing the first dataset, which is presented in Table 4. Table 4 shows the mean, standard deviation, and minimum and maximum values for each variable. It has shown that female students have a higher mean value in all variables, including final marks, CGPA, pre-requisite marks, and assessment marks. These findings align with Hyde et al.'s study [24], where females performed slightly better than males in mathematics. However, a study [25] claimed that although males and females differ very little in mathematics performance, males tend to have positive attitudes toward the subject. Additionally, a more recent study [26] found that males outperformed females in mathematics under time-pressure settings.

In terms of range, male students tended to have a wider range on each variable, except for video views and total minutes spent on videos. Therefore, in general, female students tended to perform more consistently across various variables as compared to male students. This finding is consistent with the results obtained in McSporran and Young's study [27], where female students were found to be more motivated, possess better online engagement, and manage their learning schedules more effectively than male students.

Table 3. Description of data.

Data	Descriptions	Type
Gender	Male or female students	Categorical
Final marks	The total marks obtained in a Year 2 actuarial science course taken in the semester	Numerical
CGPA	Cumulative grade point average, which is computed by dividing the total grade points from all semesters by the number of total credit hours	Numerical
Pre-requisite subject marks	The total marks of the pre-requisite subject, a subject taken by students in the previous semester	Numerical
Homework marks	The total marks obtained by students in the homework given	Numerical
Blackboard clicks	The number of clicks based on the given exercises and assessments that were produced in Blackboard	Numerical
Video views	The number of times students viewed the lecture videos	Numerical
Total minutes spent in videos	The total time students spent in the lecture videos (in minutes)	Numerical
Attendance	The attendance of students who attended workshops and tutorials (in percentage)	Numerical
Assessment marks	The total marks of a test obtained by students	Numerical

Table 4. Descriptive statistics of the data.

Data		Mean	Standard Deviation	Minimum	Maximum
Final marks		53.76			
	Male	50.89	24.96	8.00	86.00
	Female	57.63	16.26	14.50	88.00
CGPA		3.26	0.58	1.35	3.96
	Male	3.11	0.68	1.35	3.96
	Female	3.46	0.34	2.39	3.89
Pre-requisite subject marks		63.56	11.69	40.00	85.50
	Male	61.00	13.35	40.00	84.50
	Female	67.00	8.01	51.00	85.50
Homework marks		1.81	1.96	0.00	6.30
	Male	1.75	2.02	0.00	6.30
	Female	1.88	1.92	0.00	6.30
Blackboard clicks		122.00	72.47	18.00	369.00
	Male	113.77	73.18	18.00	369.00
	Female	133.08	71.40	67.00	353.00
Video views		7.57	4.80	0.00	26.00
	Male	6.51	4.05	0.00	17.00
	Female	9.00	5.41	3.00	26.00
Total minutes spent in videos		225.20	89.75	0.00	492.30
	Male	191.71	68.40	0.00	326.63
	Female	270.29	96.37	157.99	492.30
Assessment marks		7.22	2.89	1.00	14.50
	Male	6.86	3.23	1.00	12.00
	Female	7.71	2.31	4.50	14.50
Attendance		97.93	2.60	88.09	100.00
	Male	97.72	2.74	88.09	100.00
	Female	98.20	2.42	91.36	100.00

3.2.2. Correlation analysis

Correlation is a statistical technique used to assess the degree of relationship between two variables [28]. The correlation coefficient (*r* or *R*) is used to measure the degree of association of two variables. The sign of the correlation coefficient describes the direction of the correlation. A positive sign means that the two variables are moving in the same direction, i.e., when one variable increases, the other does as well. A correlation coefficient of 1 indicates a perfect relationship, 0.7 and above indicates a strong relationship, whereas a value of 0.4 to 0.6 indicates a moderate relationship, a value 0.3 and below is considered a weak relationship, and lastly 0 indicates no relationship between the variables [29].

3.2.3. Quadratic discriminant analysis (QDA)

To ensure the reliability of QDA, we will assess the underlying assumptions of QDA. The variance inflation factor (VIF) will be used to verify whether the samples are independent of each other. VIF provides a measure of multicollinearity among the independent variables, revealing

correlations between multiple independent variables. A VIF less than 5 suggests a low correlation of the variable with another variable, while a value between 5 and 10 indicates a moderate correlation. VIF values exceeding 10 indicate a high, intolerable correlation among model variables. On the other hand, we have performed Shapiro-Wilk's test to check if the assumption of multivariate normality held. The null hypothesis for this test was that the sample came from a normal distribution, where a 95% significance level was used. We also tested the equality of covariance matrices, another assumption for QDA, using Box's M test. Using the strong correlated variables identified through correlation analysis, we partitioned the dataset into 70% training data and 30% testing data to assess the model's accuracy. Finally, the prediction model was applied to identify students at risk. To determine whether a student was "at risk" or "not at risk", the final marks from the Year 2 course were considered. If a student scored below 50, they were considered "at risk", otherwise they were considered "not at risk".

3.2.4. Peer Assisted Learning Program (PALP)

The Peer Assisted Learning Program (PALP) is offered to all students who enroll in the actuarial science course. Four PALP classes were conducted throughout the semester by a senior, i.e., peer mentor. Each session was 1.5 hours and was conducted physically in the classroom. During the last class of the PALP, a survey was carried out to understand the effectiveness of the PALP using descriptive analysis and open-ended questions analysis.

3.2.5. Confusion matrix

The metrics that are commonly used in literature to measure the classification performance of a model are based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values from the confusion matrix, as shown in Table 5 [4]. With that, the effectiveness of the prediction model can be evaluated. Based on the values collected in Table 5, we applied the formula given in Table 6 [30] to calculate the classification performance of the prediction model in identifying students at risk and students not at risk. The prediction model can be considered a good model if it has high values for each metric [31].

Table 5. Confusion matrix.

	Actual Group	
Predicted Group	Not at risk	At risk
Not at risk	True Positive (TP)	False Positive (FP)
At risk	False Negative (FN)	True Negative (TN)

Table 6. Classification performance metrics and their formulas.

Metric	Formula
Accuracy	$\dfrac{TP+TN}{TP+TN+FP+FN}$
Precision	$\dfrac{TP}{TP+FP}$
Sensitivity or Recall	$\dfrac{TP}{TP+FN}$
Specificity	$\dfrac{TN}{TN+FP}$

4. Results and discussion

4.1. Correlation analysis

We applied correlation analysis to understand the variables that influenced the final marks of the Year 2 course. The results in Table 7 indicate that the final marks of the course were rather strongly correlated ($r > 0.7$) with the CGPA, prerequisite subject marks, and assessment marks. This confirms the significance of these factors in determining the overall academic success of the students. The other variables, on the other hand, show either a moderate or weak relationship with the final marks. The variables with a strong correlation will be included in the QDA to predict students at risk on the second data set.

Table 7. Correlation matrix.

Variables	Final marks	CGPA	Pre-requisite subject marks	Homework marks	Blackboard clicks	Video views	Total minutes spent in videos	Assessment marks	Attendance
Final marks	1.000								
CGPA	0.808	1.000							
Pre-requisite subject marks	0.798	0.801	1.000						
Homework marks	0.438	0.430	0.473	1.000					
Blackboard clicks	0.257	0.120	0.031	0.154	1.000				
Video views	0.427	0.276	0.216	0.193	0.603	1.000			
Total minutes spent in videos	0.217	0.161	0.126	0.145	0.281	0.627	1.000		
Assessment marks	0.703	0.685	0.702	0.406	0.246	0.309	0.074	1.000	
Attendance	0.262	0.164	0.138	0.487	0.134	0.299	0.521	0.203	1.000

4.2. Quadratic discriminant analysis to predict students at risk

Based on Table 8, it can be concluded that the three variables are independent of each other and there is no multicollinearity issue, as the VIF of each variable is less than 5. Furthermore, the observed covariance matrices for the variables are equal across groups since the p-value of Box's M test (0.062) is above 0.05. However, the null hypothesis for multivariate normality is rejected as the p-value is below 0.05 under the Shapiro-Wilk test (0.0000). The normality assumption is violated in this study due to the dichotomous nature of the robust dependent variable, meaning that this method can tolerate some deviation from the normality. Despite that the normality assumption is not met, the analysis can still be useful, according to Tabachnick and Fidell [32]. In fact, Lachenbruch [33] reviewed several studies that used discriminant analysis and found that the discriminant function performs fairly well even with non-normal data. After verifying the QDA assumptions, the prediction of students at risk using QDA proceeded with the three variables identified under correlation analysis, to form the prediction model. Upon performing the QDA with the second dataset, 54 students were identified as "not at risk" and 15 students identified as "at risk".

Table 8. Variance inflation factor scores.

CGPA	Pre-requisite subject marks	Assessment marks
3.0416	3.1823	2.1507

4.3. Survey results for the Peer Assisted Learning Program

A total of 40 students attended the PALP, which included one student who was identified as a student at risk through QDA. The students responded to the survey, which consisted of two parts. The first part had six open-ended questions with 5-point Likert scale questions ranging from 1 (poor) to 5 (superior). The results of the first part are presented in Table 9, which reports the mean and standard deviation of each close-ended question. The second part of the survey included the open-ended question, where students could provide further elaboration about the PALP.

Table 9. Responses to the close-ended questions about the PALP.

No.	Questions	Mean	Standard Deviation
1	Demonstrate good knowledge of the subject matter	4.52	0.59
2	Provide opportunities for interaction / Allow students to seek clarification	4.44	0.59
3	Explain topic content and questions clearly and effectively	4.51	0.61
4	The length of each session, i.e., 1.5 hours	4.41	0.70
5	The platform to conduct the session (physical and classroom)	4.32	0.80
6	The suitability of dates and times arranged	4.23	0.76

From Table 9, item 1 had the highest mean of 4.52, where students were asked whether the peer mentor "demonstrated good knowledge of the subject matter". On the other hand, item 6 had the lowest mean of 4.23, where students were asked about "the suitability of dates and times arranged". To understand the reasons behind the low score for item 6, some of the open-ended responses provided by students are shown below:

"Classes are too late (not able to concentrate, too tired)."
"The frequency of the PALP classes should be increased, especially when the final exam is near. Only one class was spent doing final exam questions."
"The session should be 2 hours long."

In short, students wished to have a longer PALP session, to conduct the PALP earlier, and have more PALP classes to help them prepare better for their final exam. The open-ended questions also revealed some significant keywords about the PALP, such as "good", "online", "better", "give", "questions", "answers", and "session". To support this, some selected comments given by students are presented below:

"It'll be easier for the <u>discussion</u> if the <u>questions</u> can be <u>shared</u> before the <u>session.</u>"
"It's <u>better</u> if some <u>answers</u> are <u>provided.</u>"
"Prefer an <u>online</u> <u>session</u> (could watch <u>recording</u>, clearer view of the <u>screen</u>)."
"All <u>good</u>. Thanks for the effort."

"Give more questions with answers."

To summarize the results of the open-ended comments, it appears that students preferred online PALP than a physical one. They also expressed a desire for more practice questions and for the answers to those questions to be provided. Additionally, they preferred to receive the questions before the class. Overall, the students were satisfied with the way the PALP was conducted, as long as the mentor had a good knowledge of the subject matter and was able to effectively explain the topic and address students' questions.

4.4. Assess the effectiveness of the Peer Assisted Learning Program and quadratic discriminant analysis

At the end of the semester, we compared the list of students who were correctly and incorrectly predicted as "at risk" and "not at risk" using QDA, as summarized in Table 10. Out of 15 students who were predicted as "at risk", only one attended the PALP but with poor attendance (as shown in Table 11). Therefore, they may not have benefited from it and hence failed the course. Of the remaining 14 students, who did not attend the PALP, only five managed to pass the course. On the other hand, out of the 54 students who were predicted as "not at risk", 39 of them attended the PALP and passed the course. Of the remaining 15 students who did not attend the PALP, 14 managed to pass the course, and only one did not.

Based on the analysis of the academic performance of students who attended the PALP, it was found that 97.5% of them passed the course as per Table 11. Additionally, it was noticed that students who scored at least 60 in the final marks had better attendance than those who scored below 60. However, we cannot conclude whether the PALP was an effective intervention strategy for students at risk because a majority of them failed to attend. The results are deemed to be normal since assisting students at risk requires a long-term effort and may not result in an immediate reduction in the failure rate [1]. Nonetheless, it was observed that students who attended the PALP had a higher probability of passing the course. This result is consistent with the study conducted by Cheng and Walters [23], which showed that attending the PALP for mathematics increased the likelihood of students passing the subject and completing the program.

Table 10 shows that, out of 69 students, only six students were predicted for the wrong categories, which are students no. 6, 8, 10, 12, 21, and 23, provided in the appendix. We also observed that all 10 students who failed the course (i.e., actual "at risk") were males. This is consistent with our previous observation that female students tend to perform better than male students in this course.

Using the confusion matrix formula given in Table 5, we calculated the values of accuracy, precision, sensitivity, and specificity of the prediction model. The accuracy of the model was found to be 91%, while the precision, sensitivity, and specificity were 98%, 91%, and 91%, respectively. Based on these results, we can conclude that QDA is a good prediction model for identifying students at risk for the actuarial science course. The accuracy obtained from this study (91%) is better that those reported in previous studies by Mubarak et al. (84%) and Mueen et al. (86%) [12,13].

Table 10. Summary of predicted and actual classification of students "at risk" and "not at risk" who attended the PALP.

Predicted	No. of students	Actual	No. of students	Attended PALP
Not at risk	54	Not at risk	53	39
		At risk	1	0
At risk	15	Not at risk	5	0
		At risk	10	1
	69		69	40

Table 11. Final marks and attendance for students who attended the PALP for the subject.

Final marks	No. of students	Percentage (%)	Average attendance (%)
≥ 70	19	47.5	65.8
60 - 70	12	30.0	33.3
50 - 60	6	15.0	62.5
40 - 50	2	5.0	25.0
< 40	1	2.5	25.0
	40	100	

5. Conclusions

To summarize, we successfully developed a QDA prediction model to identify students at risk for an actuarial science course with high levels of accuracy, precision, sensitivity, and specificity. Our prediction model relied solely on students' learning behaviors related to academic performance, i.e., CGPA, pre-requisite subject marks, and assessment marks. While we cannot conclude that the PALP is an effective intervention strategy for students at risk, our results shows that students who attended had a significantly higher chance of passing the course. Moving forward, our focus will be on finding ways to encourage more students at risk to attend the PALP. In addition, we will consider the feedback we have received from students to improve the program.

In this study, we observed that all 10 students who failed the course were male. This finding can be investigated further since previous studies have suggested that male students tend to perform better than female students in mathematics. To validate these results, we need a larger sample size. Additionally, this raises questions about whether we should pay extra attention to male students or create a customized intervention program for them. These are areas that require further investigation in future studies.

Use of AI tools declaration

The authors would like to disclose that an AI tool was utilized in the development of this paper. The primary AI tool used was Grammarly, which was used to assist in improving writing, providing suggestions, and enhancing the overall quality of the written paper.

Acknowledgments

Conflict of interest

The authors declare that there are no conflicts of interest in this paper.

Ethics declaration

The authors declare that the ethics committee approval was waived for the study.

References

1. Lewis, R. and McCann, T., Teaching "at risk" students: Meeting their needs, in *International Handbook of Research on Teachers and Teaching*, Saha, L.J., Dworkin, A.G, Ed. 2009. Springer International Handbooks of Education.

2. Alyahyan, E. and Düştegör, D., Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 2020, 17(1): 3. https://doi.org/10.1186/s41239-020-0177-7

3. Rienties, B., Nguyen, Q., Holmes, W. and Reedy, K., A review of ten years of implementation and research in aligning learning design with learning analytics at the Open University UK. *Journal of Interaction Design and Architecture*, 2017, 33: 134–154.

4. Choi, S.P.M., Lam, S.S., Li, K.C. and Wong, B.T.M., Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Educational Technology & Society*, 2018, 21(2): 273–290.

5. Wilkinsona, K., McNamaraa, I., Wilsonb, D. and Riggsa, K., Using learning analytics to evaluate course design and student behaviour in an online wine business course. *International Journal of Innovation in Science and Mathematics Education*, 2019, 27(4): 97–108. https://doi.org/10.30722/IJISME.27.04.008

6. Abdul Jalil, N. and Wong Ei Leen, M., Learning analytics in higher education: The student expectations of learning analytics. *Proceedings of the 2021 5th International Conference on Education and E-Learning* (ICEEL '21). New York, NY, USA, 2021, 249–254. https://doi.org/10.1145/3502434.3502463

7. Arnold, K.E., Signals: Applying academic analytics. *Educause Quarterly*, 2010, 33(1): n1.

8. Ramli, I.S., Maat, S.M. and Khalid, F., Learning analytics in mathematics: A systematic review. *International Journal of Academic Research in Progressive Education and Development*, 2019, 8(4): 436–449. http://dx.doi.org/10.6007/IJARPED/v8-i4/6563

9. Akçapınar, G., Altun, A. and Aşkar, P., Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 2019, 16(1): 1–20. https://doi.org/10.1186/s41239-019-0172-z

10. Campbell, J.P. and Oblinger, D.G., Academic analytics. *EDUCAUSE Review*, 2007, 42(4): 40–57.

11. Finch, H. and Schneider, M.K., Classification Accuracy of Neural Networks vs. Discriminant Analysis, Logistic Regression, and Classification and Regression Trees. *Methodology*, 2007, 3(2): 47–57. https://doi.org/10.1027/1614-2241.3.2.47

12. Mubarak, A.A., Cao, H. and Zhang, W., Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, 2020, 30(8): 1414–1433. https://doi.org/10.1080/10494820.2020.1727529

13. Mueen, A., Zafar, B. and Manzoor, U., Modeling and predicting students' academic performance using data mining techniques. *International Journal Modern Education and Computer Science*, 2016, 11: 36–42. https://doi.org/10.5815/ijmecs.2016.11.05

14. Ha, D.H., Nguyen, P.T., Costache, R., Al-Ansari, N., Van Phong, T., Nguyen, H.D., et al., Quadratic Discriminant Analysis based ensemble machine learning models for groundwater

potential modeling and mapping. *Water Resources Management*, 2021, 35(13): 4415–4433. https://doi.org/10.1007/s11269-021-02957-6

15. McLachlan, G.J., *Discriminant Analysis and Statistical Pattern Recognition.* 2nd ed. 2004, Hoboken, NJ, USA: Wiley.

16. Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. 2009, New York, NY, USA: Springer.

17. Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., et al., A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Australia, 2015, 1909–1918. https://doi.org/10.1145/2783258.2788620

18. Yang, S., Lu, O., Huang, A., Huang, J., Ogata, H. and Lin, A., Predicting students' academic performance using multiple linear regression and principal component analysis. *Journal of Information Processing*, 2018, 26: 170–176. https://doi.org/10.2197/ipsjjip.26.170

19. Shah, R.K. and Barkas, L.A., Analysing the impact of e-learning technology on students' engagement, attendance and performance. *Research in Learning Technology*, 2018, 26. https://doi.org/10.25304/rlt.v26.2070

20. Stewart, M., Stott, T. and Nuttall, A-M., Student engagement patterns over the duration of level 1 and level 3 Geography modules: influences on student attendance, performance and use of online resources. *Journal of Geography in Higher Education*, 2011, 35(1): 47–65. https://doi.org/10.1080/03098265.2010.498880

21. Nepal, R. and Rogerson, A.M., From theory to practice of promoting student engagement in business and law-related disciplines: the case of undergraduate economics education. *Education Sciences*, 2020, 10(8): 205. https://doi.org/10.3390/educsci10080205

22. Makala, Q., Peer-assisted learning programme: Supporting students in high-risk subjects at the mechanical engineering department at Walter Sisulu University. *Journal of Student Affairs in Africa*, 2017, 5(2): 17–31. https://doi.org/10.24085/jsaa.v5i2.2700

23. Cheng, D. and Walters, M., Peer-assisted learning in mathematics: An observational study of student success. *Journal of Peer Learning*, 2009, 2: 23–39.

24. Hyde, J.S., Fennema, E. and Lamon. S.J., Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, 1990, 107(2): 139–155. https://doi.org/10.1037/0033-2909.107.2.139

25. Else-Quest, N.M., Hyde, J.S. and Linn, M.C., Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, 2012, 136(1): 103–127. https://doi.org/10.1037/a0018053

26. Arias, O., Canals, C., Mizala, A. and Meneses, F., Gender gaps in Mathematics and Language: The bias of competitive achievement tests. *PLoS ONE*, 2023, 18(3): e0283384. https://doi.org/10.1371/journal.pone.0283384

27. McSporran, M. and Young, S., Does gender matter in online learning? *Research in Learning Technology*, 2001, 9(2): 3–15. https://doi.org/10.1080/0968776010090202

28. Mukaka, M., Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 2012, 24(3): 69–71.

29. Dancey, C.P. and Reidy, J., *Statistics Without Maths for Psychology: Using SPSS for Windows*. 4th ed. 2008. Prentice Hall.

30. Karalar, H., Kapucu, C. and Gürüler, H., Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *International Journal of Educational Technology in Higher Education*, 2021, 18(1): 63. https://doi.org/10.1186/s41239-

021-00300-y

31. Tsakirtzis, S. and Georgakopoulos, I., Developing a risk model to identify factors which critically affect secondary school students' performance in mathematics. *Journal for the Mathematics Education and Teaching Practices*, 2020, 1(2): 63–72.

32. Tabachnick, B.G. and Fidell, L.S., *Using Multivariate Statistics*. 6th ed. 2012, Boston, MA, USA: Person Education.

33. Lachenbruch, P.A., Discriminant Analysis. *Biometrics*, 1979, 35: 69–85.

Author's biography

Chan Lay Guat is a Senior Lecturer of the School of Mathematical Sciences at Sunway University, Malaysia. Her research areas are actuarial science and mathematics education.

Ng Qian Yun graduated from the BSc (Hons) Actuarial Studies program at Sunway University. Her research areas are actuarial science and mathematics education.