Using Experiential Learning Activities in Simulation Games to Predict Students' Scores

Nang Laik Ma, Singapore University of Social Sciences, Singapore
Ivy Sook May Chia, Singapore University of Social Sciences, Singapore

The Asian Conference on Education 2024
Official Conference Proceedings

Abstract
Analytical and problem-solving skills are crucial for thriving in the workplace instead of mere content knowledge. To better prepare our undergraduates for entry into the workforce in this tumultuous time, Experiential Learning Theory (ELT) has been employed in the business programs. A cloud-based simulation game called MonsoonSIM has been deployed in one of the introductory courses in the business school. The simulation game aims to allow students to explore a broad spectrum of business processes ranging from retail, e-commerce, wholesales, manufacturing, procurement, human resources planning, forecasting, accounting, and finance. Through experiential learning and collaboration with teammates via an online portal, students are encouraged to deepen their understanding by playing the game online. In this paper, we aim to analyze the students' activities in the simulation games and use it as a proxy to measure their engagement level and take preemptive action to harness students' problem-solving and data analysis skills. The authors have collected hundreds of students' data from two semesters and used anonymized students' activities and the pre-class quiz results to predict the student's final scores for the course. The regression model is proposed using input as the students' activities and one of the pre-class quizzes to predict the students' final scores. The model accuracy rate is measured using Mean Absolute Percentage Error (MAPE), which is less than 10% and is an excellent predictive model. It helps the educator to analyse the student's performance early in the course and improve their overall learning experience.

Keywords: Experiential Learning Activities, Simulation, Predictive Model

Introduction

Every business school undergraduate must take our university's Business Skills and Management course. The course objective is to help students develop problem-discovery and problem-solving skills systematically using spreadsheets Excel. The course is the first course they take; thus, the successful completion of the course is vital for the students. The course focuses on experiential learning and problem-solving to prepare students to cultivate self-directed learning and equip them with the necessary skills to survive in challenging real-world situations.

The students come from various academic backgrounds, most of them from polytechnics, and thirty percent of them are from junior colleges. Many of them may need to gain prior knowledge in business to enable them to follow in class. In 2017, we first introduced a business simulation game called MonsoonSIM in the course to help students understand the overview of running a business. There are more than ten departments in the game, which range from accounting and finance, marketing, logistics, retail, wholesale, e-commerce, production and warehouse, manufacturing, service, and human resource departments. It allows students to explore the departments through experiential learning. Over the years, students find the simulation game engaging and enhancing their learning and understanding.

In our course, students learn to develop business models using Spreadsheets from scratch, making valid assumptions. The assessment criteria are based on three pre-class quizzes, individual assignments, and a class test. If a student fails the course, it could significantly impact their confidence and future studies. This underscores the importance of early intervention in identifying students who may need assistance in the course before it's too late, a key focus of our research.

The course is taught over twelve weeks of class over a semester. Six weeks before the start of the class, students will have access to the online learning portal, which they can self-study using a study guide, e-textbook, and PowerPoint slide. In the first lesson, students will be assigned randomly into groups of five, and they will play the simulation game for two hours. Recently, we found out that the learner activities report from the simulation game can be used as a proxy for their engagement and commitment to the course. Thus, in this paper, we explore using the learner activities and the first-pre-class quiz, which happens in the first week, to predict the student's final score. The outcome of this research will assist the lecturers in identifying students who may need assistance in the course before it is too late to do anything.

The rest of the paper is organized as follows. In section two, we will do a literature review to look at the predictive model to improve students' performance and outcomes in education. Next, we will discuss how to collect the data and share some preliminary data analysis and insight. In section four, we will develop predictive models using the input data such as learner activities count, pre-class quiz score, and final score. Finally, we discussed the model's accuracy and recommended actions required using the insights to improve the student's academic performance, highlighting the potential impact of our research on student outcomes.

Literature Review

Many researchers, such as Aldowah et al. (2019), and Chiappe & Rodriguez (2017), use the students' data from the learning management system to improve the students' academic achievement.

Barrows (1996) explained the importance of fully supporting faculty in developing a new curriculum based on a problem-based learning (PBL) approach. The faculty wanted to use PBL to see its impacts on students' learning capability and independent thinking.

Other authors, Ma & Chia (2020), developed a new learning analytics course for the master-degree program. The course mainly focuses on problem-based learning (PBL) to solve real-world problems in the classroom environment and has received good end-of-course evaluations from students. Ma & Chia (2023) developed a case study for the learning analytics course to predict the students' cumulative gross point average (CGPA) based on the five courses. Three predictive models, decision tree, regression, and neural network, have been developed, and the model performance based on mean absolute error (MAE) showed that the regression model yields the slightest error. Thus, it is the champion model.

Students are leaving digital traces online. Some of them, such as the number of pages read, days interacted, time spent, and the number of highlights, bookmarks, and notes, can be used as a proxy to determine the student's engagement level. Junco and Clem (2015), the authors use digital student information to identify at-risk students using digital course reading and engagement. Additional data sources include previous GPA, course grades, and demographic information. The result showed that the engagement and number of days students spent on reading were strong predictors of student performance. The developed system can help educators identify weaker students and provide additional coaching sessions to improve their academic results.

Data Preparation

Student ID is used as the key to match the student's score as well as the student's activity count. We have masked the students' information and created a new student ID to identify the students. There are nearly hundreds of students’ records from the past semester. Students' ID is categorical; learner activity count is a numerical number greater or equal to zero. Pre-class quiz score is a numerical number between zero and 100. The final target variable is set as the final score for the course, which is numerical between zero and 100.

Table 1: Student Data

Description	**Data Field**
Student ID	Categorical
Learner activity count	The number of activities done by students in the first game.
Pre-class quiz score	Quiz score (0 - 100)
Final score (Target)	Final score (0 - 100)

Next, let us explore the descriptive statistics of the input variable. The mean activity count is 23.2, but the standard deviation is 19, meaning there is a high variation in the students' activity levels during the game. The minimum number is zero, and the maximum activity count is 85. However, the mean pre-class quiz score is 81.4, considered high. The average final score of the course is 71.2, the mode is 75.1, and the standard deviation is 8.33.

Table 2: Summary Statistics of Students' Data

	Learner activity count	**Pre-class quiz score**	**Final score**
Mean	23.2083	81.4063	71.1906
Standard Error	1.9447	1.6419	0.8508
Median	19.0000	85.0000	72.1250
Mode	20.0000	90.0000	75.1000
Standard Deviation	19.0539	16.0870	8.3365
Sample Variance	363.0509	258.7911	69.4964
Kurtosis	2.3092	12.2507	1.6563
Skewness	1.6080	-2.8724	-1.1333
Minimum	0	0	45.1
Maximum	85	100	88.9

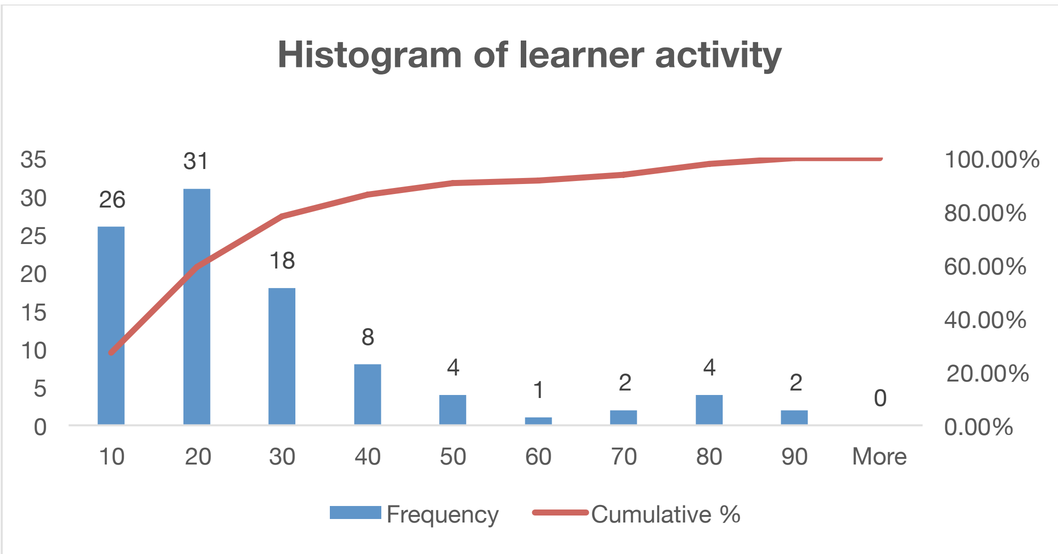

Figure 1: Histogram of Learner Activity

Figure 1 shows that 78% of the students contributed to the learning activity for 10 to 30 intervals. Less than 10% of the students have a learning activity above 50.

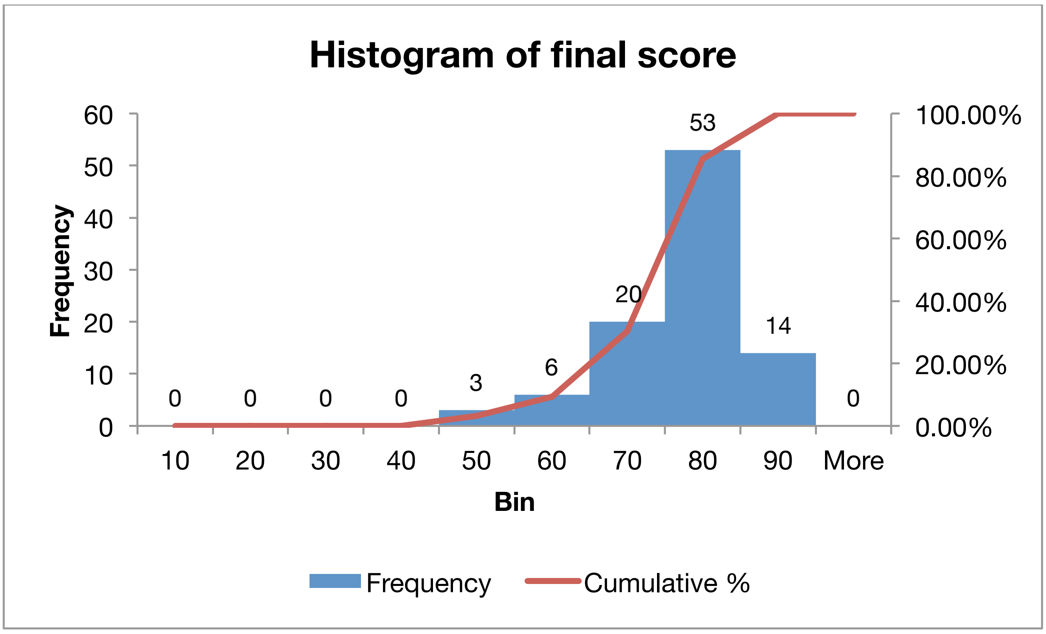

Figure 2: Histogram of the Final Score

Figure 1 shows the histogram of learner activity left-skewed, with the highest frequency between 10 and 20. However, the histogram of the final score, Figure 2, shows that the histogram is right-skewed, with the highest frequency around 70 to 80; 85% of the students scored less than 80, and only 15% scored above 80.

Regression Models and Computational Results

In this section, the authors showed how to develop the predictive model using input variables such as learner activity count. A regression model is chosen because it is easier to understand the relationship between the dependent variables (X's) and independent variable Y. It is one of the most popular predictive models deployed in real business scenarios.

Regression is a statistical model that finds the relationship between the independent variable Y and one or more dependent or explanatory variables X. The assumption is that a linear relationship exists between the dependent variables X's and the independent variable Y.

Let i be the students, i = 1 to N
Let Y_i be the final score of student i
Let $X1_i$ be the learner activity count of student i
Let $\hat{Y}_i$ be the predicted final score of student i

Model 1 is built only on the learner activity, where $\hat{Y}$ is the predicted final score. Using regression analysis, we can get the linear equation,

$$\hat{Y} = 0.0177\ X1 + 70.78 \tag{1}$$

We can use the equation to compute the predicted final score for all the students.

If X1 is 60, the predicted score is calculated using the equation $\hat{Y} = 0.0177 * (60) + 70.78 = 71.84$. We can then compute the actual final score and the absolute percentage error. Assuming the actual score is 75, the absolute percentage error in this case is 4%.

$$\text{Mean absolute percentage error (MAPE)} = \frac{\sum_{i=1}^{i=n} \frac{|Y_i - \hat{Y}_i|}{Y_i}}{n} * 100\% \quad (2)$$

Using the above formula, we can compute the absolute percentage error for each student and calculate the average absolute percentage error. Model 1 yields a MAPE of 8.47%, which shows that students' activity counts can be used as a predictor for the final score. However, there is an issue with the model, as the student's activity count is between 0 and 85; using the equation, the minimum score for the students with no activity will be 70.78, which is the y-intercept. But it is not true as the students can score less than 70. Based on the historical data, the percentage of students who score less than 70 is about 30%. Thus, we need to add more variables in the next model.

We want to develop a predictive model to predict the final score using the learning activities during the simulation and pre-class quiz in the first lesson. Pre-class quizzes and simulation games are conducted in the same week during the first lesson. If we can use it to predict the students' final scores, we can preempt students who do not do well in these two components to put in more effort and improve the course outcome.

Let w_1, w_2 be the weight assigned to learner activity count and pre-class quiz accordingly.

Let i be the students, i = 1, 2, … to N
Let Y_i be the final score of student i
Let W_i be the weighted score of student i
Let $X1_i$ be the learner activity count of student i
Let $X2_i$ be the pre-class quiz score of student i
Let $\hat{Y}_i$ be the predicted final score of student i

Model 2 is built only on the weighted score of the learner activity and pre-class quiz, where $\hat{Y}$ is the predicted final score.

We initially set an equal weightage of 50% for each learner activity count and pre-class quiz score.

$$W_i = w_1 * X1_i + w_2 * X2_i \quad (3)$$

The general regression line to predict the student's score is $\hat{Y}$ = intercept + slope * weighted score. We can get the linear equation,

$$\hat{Y} = 0.2267 \text{ W} + 59.33 \quad (4)$$

Using regression analysis, we can use the equation to compute the predicted final score for all the students. Next, we want to add the mean absolute error above in equation (2).

Using equation (2), we can compute the absolute percentage error for each student and calculate the average absolute percentage error. Model 2 yields a MAPE of 7.96%, which shows that students' activity counts and pre-class quizzes can be used as predictors for the final score, and the error is smaller than just using one variable. We also want to find the optimal weightage for the two components, which minimizes the MAPE. The only constraint added to the model is that the sum of weight equals 1.

$$w_1 + w_2 = 1 \tag{5}$$

Using the Excel solver option, we can compute the optimal weight of 30% for the learner activity and 70% for the pre-class quiz. The minimum MAPE is 7.80%.

Figure 3 shows that by varying the weight for w1 from 10% to 90%, MAPE reduces from 7.93% to 7.8% as the minimum MAPE when w1 equals 30%. After which, MAPE increases to 8.43% when w1 = 90%.

Thus, the optimal weight for w_1, which is the weight for the learner activity, is 30% and 70% for w_2 for the pre-class quiz. This is the weight of the pre-class quiz and will yield the minimum MAPE of 7.80%.

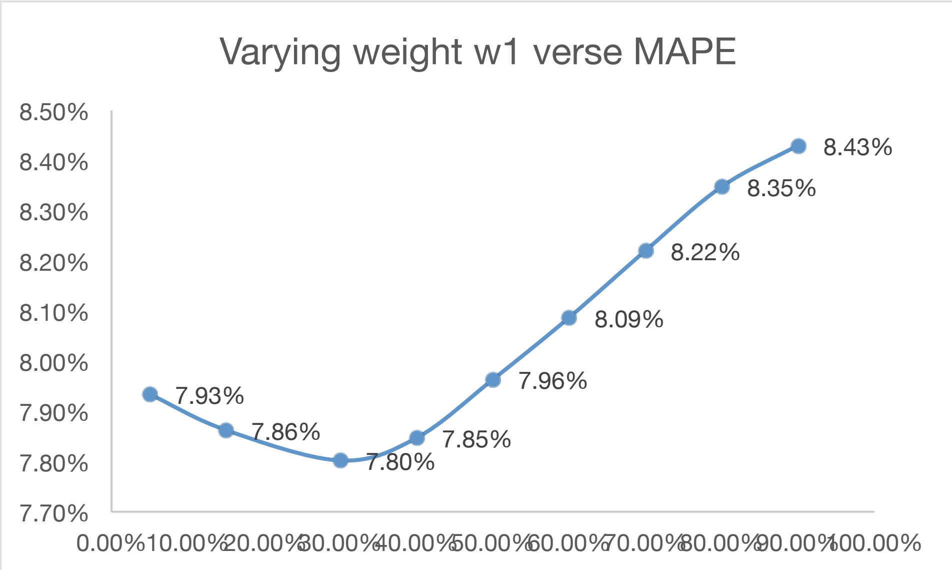

Figure 3: Varying Weightage for the Learner Activity Count and MAPE

Hypothesis Testing

a. We are only using the learner activities.

We are using hypothesis testing to determine whether there is any difference between the predicted and observed scores, where the predicted score is only based on the simulation learner activity.

Let d_i be the difference between the observed score and the predicted score
$d_i = Y_i - \hat{Y}_i$ where is the i = 1, 2, .., N

$H_0: \mu_d = 0$
$H_1: \mu_d \neq 0$

At α=0.05, we use a paired t-test as the data is dependent. The score follows a normal distribution, as we observed from the histogram.

Using Excel→ Descriptive statistics → t-test. We get the result as shown below.

t-Test: Paired Two Sample for Means

	Predicted score	*Final*
Mean	71.191	71.191
Variance	0.114	69.496
Observations	96	96
Pearson Correlation	0.04046	
Hypothesized Mean Difference	0	
df	95	
t Stat	1.3320E-14	
P(T<=t) one-tail	0.5	
t Critical one-tail	1.6611	
P(T<=t) two-tail	1.00	
t Critical two-tail	1.9853	

The p-value is nearly 1, which is more than α value of 0.05, thus we cannot reject H_0, thus accepting H_0. This means the difference between the predicted and actual scores is zero.

b. Using learner activity and pre-class quiz

We are using hypothesis testing to check for any difference between the predicted and observed scores, where the predicted score is the weighted score based on the simulation learner activity and pre-class quiz.

Let d_i be the difference between the observed score and the weighted predicted score
d_i =$Y_i - \hat{Y}_i$ where is the i = 1, 2, .., N

H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$

At α=0.05, we use a paired t-test as the data is dependent. The score follows a normal distribution, as we observed from the histogram.

Using Excel→ Descriptive statistics → t-test. We get the result as shown below.

t-Test: Paired Two Sample for Means

	Predicted score	*Final*
Mean	71.1906	71.1906
Variance	8.3071	69.4964
Observations	96	96
Pearson Correlation	0.34574	
Hypothesized Mean Difference	0	
df	95	
t Stat	1.122E-14	
P(T<=t) one-tail	0.5	
t Critical one-tail	1.6611	
P(T<=t) two-tail	1	
t Critical two-tail	1.985	

The p-value is nearly 1, which is more than α value of 0.05. Thus, we can't reject H_0 and cannot accept H_1. We accept H_0 such that $\mu_d = 0$.

Thus, there is no statistical difference between the weighted predicted and observed scores.

Conclusion

In conclusion, we developed regression models to use simulation game learner activity and pre-class quiz scores as predictors to predict the student's scores for the first course they took at our university. The models developed showed that the mean absolute percentage error MAPE is only 7.8% and can be used as an early indicator to estimate the students' scores. Those students who score less than 60 in the predicted score might be at risk of performing poorly in the course. Thus, as educators, we can take preemptive action to pay more attention to these students, provide guidance, and conduct extra lessons to enhance the student learning experiences and outcomes.

References

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37(37), 13–49.

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, 1996(68), 3–12.

Chiappe, A., & Rodríguez, L. P. (2017). Learning Analytics in 21st century education: a review. *Ensaio: Avaliação E Políticas Públicas Em Educação*, 25(97), 971–991.

Dziuban, C., Moskal, P., Cavanagh, T., & Watts, A. (2012). Analytics that Inform the University: Using Data You Already Have. *Journal of Asynchronous Learning Networks*. 16(3), 21-38.

Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education*, 27, 54–63.

Ma, NL., & Chia, I. (2020). Designing Learning Analytics for Effective Learning Through Problem-Based Learning. *Journal of Education and Practice*. Vol 11, No 18 (2020).

Ma, NL., & Chia, I. (2023). Using education analytics to predict at-risk students in the university: Case study. *Conference Proceedings of Decision Science Institutes 2023, USA Atlanta.*

Contact email: nlma@suss.edu.sg