

PAPER • OPEN ACCESS

Student Learning Progress as Predictor for Graduate Employability Performance

To cite this article: Wan Nor Afiah Wan Othman and Aziman Abdullah 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **769** 012019

View the [article online](#) for updates and enhancements.

Student Learning Progress as Predictor for Graduate Employability Performance

Wan Nor Afiqah Wan Othman and Aziman Abdullah

Faculty of Computing (FKOM), College of Computing and Applied Sciences,
Universiti Malaysia Pahang, Kuantan, Pahang, Malaysia

Email: wnor_afiqah93@yahoo.com, aziman@ump.edu.my

Abstract. Graduate employability is a major concern for higher education industry. There is a lack of research on the use of program learning outcomes (PLO) data to predict graduate employability performance especially on the duration they get employed. Therefore, our motivation in this study is to investigate how PLO data can be used to predict graduate employability performance. This study adopted quantitative analysis as a research method by using Simple Linear Regression to measure the highest correlation and significance values between learning progress and duration graduate to get employed. The PLO data from all semesters were segmented into four-time segments: 1st SEM, MID SEM, Pre-LI and LI. The slope value of linear model from time series analysis of four-time segments is used as a value to determine the performance of student learning progress. 47 responses (22% response rate) from 216 graduates who completed their study from Faculty of Computing, Universiti Malaysia Pahang in 2018 has been received as a case study. We found that learning progress from PLO 3 and PLO 6 which are '*Social Skills and Responsibilities*' and '*Problem Solving and Scientific Skills*' respectively, show significant values on the duration to get employed. This study highlights student learning progress is potential to be used as a predictor for graduate employability performance.

1. Introduction

In the era of Industrial Revolution 4.0, there is more and more higher education institution putting their concern on the issue of graduate employability. Graduate employability is very much influence with individual competencies in the field as well as generic competencies such as problem solving, communication and other skills most researchers refer as soft skills.

Many literature studies on student performance using data mining techniques. Although the techniques are capable to predict the academic performance, it is, however, depending on what data or attributes are been used. Most study using attributes like CGPA and psychometric data to predict academic performance [1]. These attributes mostly are not much helpful for the institution to predict graduate employability based on their skills or competencies level. Therefore, our motivation in this study is to investigate how learning outcomes data can be used to predict graduate employability performance. This prediction is important for higher education institution to improvise teaching and learning strategy through interventions approach. This is well supported by EDUCAUSE study [2] where most institution (96%) cited that improving student outcomes was a goal for their student success studies.

This study aims to identify the feasibility of student learning progress based on learning outcomes to predict graduate employability performance. To facilitate this study, the key research questions are:



RQ1: How to develop a predictive model for student learning progress?

RQ2: Does learning progress significance or influence graduate employability performance?

2. Literature review

2.1. Learning outcomes

Learning outcomes can be referred to as a student's competencies that are shown after the completion of their studies [3–8]. The international trends of learning outcomes in higher education are based on 'student-centred' approach [9]. This approach points out the outputs related to student competencies. This learning outcomes stated in the program and course information are part of the important component in outcomes-based education (OBE). The program learning outcomes (PLO) is a foundational requirement in the OBE system [10]. In Malaysia Qualification Framework (MQF) version 1.0, there are eight standards of program learning outcomes (PLO) that any institution in Malaysia minimally must comply in order to be accredited. Table 1 shows the eight stipulated domains in PLO.

Table 1. The eight domains of program learning outcomes (PLO)

Program Learning Outcomes (PLOs)	Attributes
PLO-1	Knowledge
PLO-2	Practical Skills
PLO-3	Social Skills and Responsibilities
PLO-4	Values, Attitudes and Professionalism
PLO-5	Communication, Leadership and Team skills
PLO-6	Problem Solving and Scientific Skills
PLO-7	Information Management and Lifelong Learning Skills
PLO-8	Managerial and Entrepreneurial Skills

2.2. Learning progress

A potential way to visualize the students regulate effort is by using learning progress. The progress illustrates how much the student has learned a specific skill during their learning over time [11]. The continuous feedback to students is important as they can monitor and control their learning progress. Winne & Hadwin highlighted the importance of data that influence the learning progress: *without reliable, revealing and relevant data that support learners to make valid inferences about how they control and monitor their learning, learners will be handicapped* [12].

Molenaar, Horvers & Baker applied moment-by-moment learning progress to investigate the students' self-regulated learning [13]. This study concluded that learning progress has a relationship with students' accuracy and learning, but not with effort. This finding is consistent with another study [11], where the effort is not showed directly, but by visualizing the student recent progress can make them connect the progress with their recent effort. Another literature [14] used data mining to predict and solve the problem on student performance estimation, student progress, and student potentials. They used student progress indicators and attribute causal relationship predictor where it shows up the factor that affects the student performance. The results indicate the proposed tools give correct and accurate results as well as perform a better understanding on student progress. The study, however, did not address the graduate employability performance.

2.3. Prediction on Graduate Employability

Study by [15] adopted linear regression to look into the predictors of graduates' employability based on graduates' competencies. The study used a questionnaire for data collection and found that value-related graduates' competencies are important for employability. However, the study did not predict the duration of graduate to be employed and does not use learning outcomes data. Based on the related

works, this study also adopts the simple linear regression method in predicting graduate employability through student learning progress based on the program learning outcomes data.

2.4. Research Gap

Although there are many literature studies on learning outcomes, studies on learning outcomes that addressing graduate employability performance are still lacking particularly in predicting the duration to get employed. It is found that there are only 38 literatures on graduate employability (title) that associate with learning outcomes (keyword) based on our simple search (1:05 pm; 14/9/2019) from google scholar database. Figure 1 shows the relative comparison in Venn diagram the total number of literatures related to '*learning outcome*' (LO) and '*graduate employability*' (GE). In this study, those articles will not be reviewed since our study are based on the case study in Universiti Malaysia Pahang (UMP) which we can confirm is the first attempt that UMP conducting a study to predict graduate duration based on learning outcomes data.

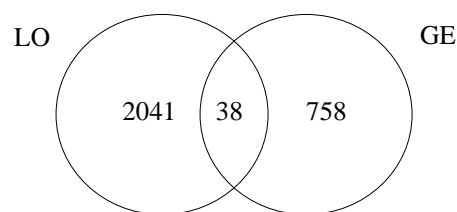


Figure 1. Total number of articles based on the keywords from googles scholar

3. Methodology

Figure 2 shows the research model adopted in this study. Predictive Analytics Process adapted from the actual version (refer [16]) was used.

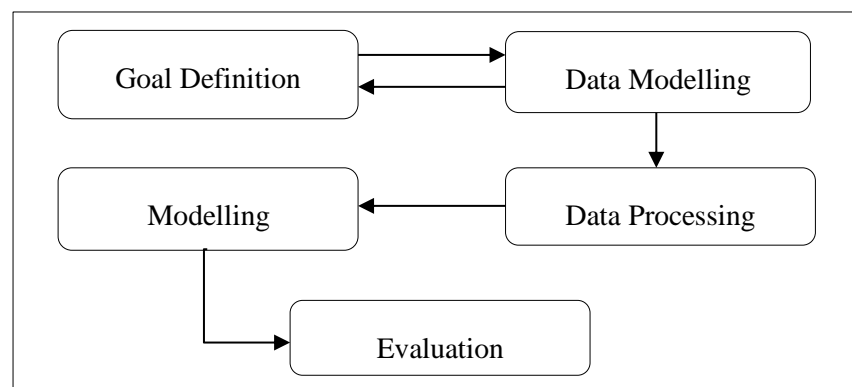


Figure 2. Predictive analytic process – modified version from Poornima & Pushpalatha (2018)

3.1. Phase 1: Goal definition

The goal defined for this study is to identify which competency through learning progress that affects the duration of graduate to employed. This part is very important that will determine the required data model useful for the predictive model. In this study, the goal for the predictive model is based on the mission of National Graduate Employability Blueprint 2012-2017 [17] which to ensure 75% of the graduate be employed within six months in their field of study.

3.2. Phase 2: Data modelling

There are two data sources used in this study which are an institutional academic database (PTMK UMP) and online feedback from the graduate. Table 2 illustrates a summary of the data model adopted in this study.

Table 2. Data model

Data	Attributes	Source
Program Learning Outcomes Attainment	Student ID , Enrolled programmed, Semester, Faculty and all eight program learning outcomes attainment scores.	Institutional Academic Database (permission granted).
Graduate Employability	Student ID and dates of first employment after graduate.	Online feedback.

3.3. Phase 3: Data processing

Time duration is a data value that dynamically changes over the time itself. We only asked the graduate to response simply the first date that they are get employed for the first time after graduate. Based on the given date, we performed data processing to find the duration by calculating the number of days between two known dates based on equation (1).

$$\text{Duration (day)} = \text{date of employment} - \text{date of graduation} \quad (1)$$

Table 3 shows the structure of the data how duration of employed graduate is been produced. This study adopts a common spreadsheet application with the existing formula in calculating the number of days between two dates.

Table 3. Data discretion of employment within six months attributes

Student ID	Date of graduation	Date of Employment	Duration (Number of Days)
Student 1	2018-07-25	2018-08-06	12
Student 2	2018-07-25	2018-11-26	124

Another data used to develop our model is student learning outcomes data. There are variations of the total mark for each PLO been assessed in each semester. We transform the data by producing the PLO attainment score by calculating the ratio of student score relative to total mark within scales 0.0 to 1.0 based on equation (2). For example, if a student got a score of 80 marks over 100 total marks for PLO1, then the ratio will be 0.8.

$$\text{PLO Attainment Score} = \frac{\text{Total Student Mark}}{\text{Total Full Mark}} \quad (2)$$

The achievement of learning outcomes may not sufficient to reflect the capability of a student in learning since the learning process itself requires a progressing approach [18]. To measure the learning progress, this study used the slope value of linear equation based on learning outcomes data projected with time-series analysis. To produce time-series analysis, the data of graduate learning outcomes have been segmented into four-time segments: 1st SEM, MID SEM, Pre-LI and LI as shown in figure 3. The 1st SEM is referring to the first semester when student admitted or enrolled in the program. The MID SEM is the maximum score of PLO from the second semester and the last two semesters before they went to internship. While Pre-LI and LI are referring to the last semester graduate or student studied in university and the semester when they were doing industrial internship. The reason for this segmentation is to standardize the data model since there are variations in term of the total number of semester graduates completed their studies.

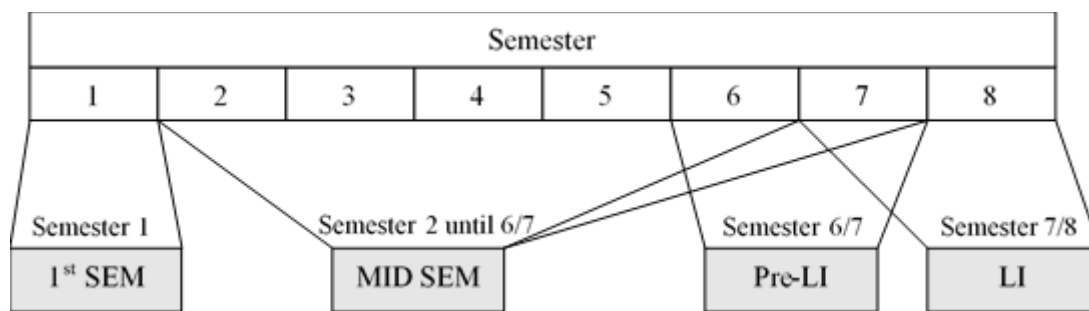


Figure 3. Data discretion of semester 1 until 8 into four groups

Once all the data for each graduate been segmented into four-time segments, then time series analysis is been conducted as illustrated in figure 4. With this dataset, we calculate the value of slope from the linear equation model as a value that represents student learning progress of each PLO for each student.

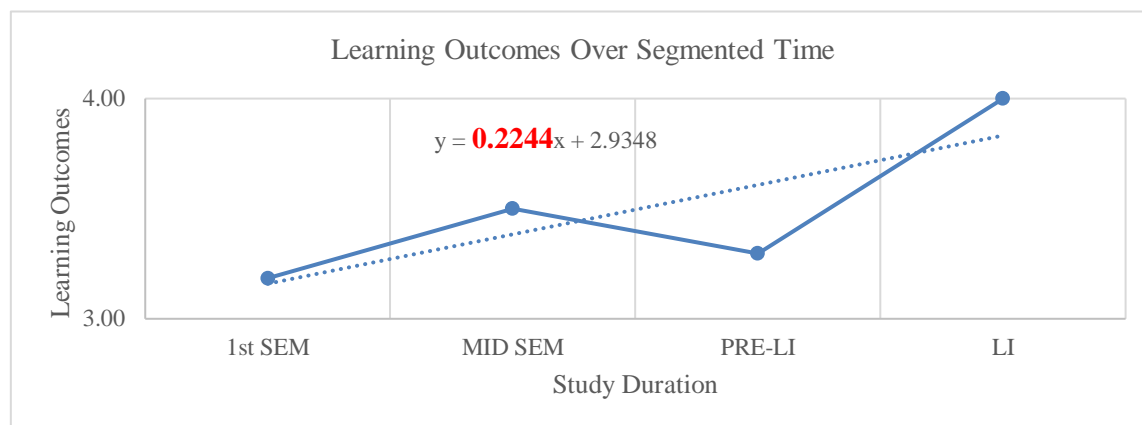


Figure 4. Learning progress model for one student

3.4. Phase 4: Modelling

Modelling phase in which training and testing data into statistical methods will be performed to create the model. The model of candidates, selection and validation are included in this phase. In this study, a simple linear regression (SLR) method will be used to build the model. The SLR is used because it is simple, direct, effective and easy to understand [19].

Model of candidates is a process to determine the correlation coefficient (r) and coefficient of determination (R^2) for the relationship between y - dependent variable (duration of employment within six months after graduation) and x - independent variable (learning progress) of each candidate. Pearson Product Moment Correlation (r) [20] is used. The significance testing (p -value approach) is used to make conclusions of the hypothesis whether the data support or rejects the null hypothesis (H_0). The p -value is compared to a significance level (α). In short, there are:

$p\text{-value} < \alpha \Rightarrow \text{reject } H_0 \Rightarrow \text{accept } H_a$

$p\text{-value} \geq \alpha \Rightarrow \text{fail to reject } H_0$

where H_a is expecting the graduate being employed in their related field of study within six months upon graduation. While H_0 is considered as vice versa from H_a .

Model selection process produces the best SLR model to be select among the eight PLOs. To obtain the best model to use for the testing and predicting upcoming data sets, the statistical method should be carried out, then the steps given need to be followed:

Step 1: Determine the R^2 and r . If positive value, r will be rejected.

Step 2: Determine the p -value where the result of attributes is significant relationship.

The process of model validation aims to identify prediction accuracy performance. It is implemented when the selected model is tested using the same dataset. The model validation is conducted in order to prove the attributes used in this study able to build the correct models. The accuracy performance of the model is validated based on the difference between the predicted value and actual value relative to the acceptable range of employed duration (6 months). The error rate of the model is defined as:

$$\text{Error rate (\%)} = \frac{\text{Predict} - \text{Actual}}{\text{Goal}} \times 100 \quad (3)$$

where,

Predict = Predicted duration using SLR model

Actual = Actual duration from raw data

Goal = 6 months or approximately 182 days

3.5. Phase 5: Evaluation

The evaluation phase is to measure the performance of linear model in UMP as a case study. The frequency analysis of error rate will be conducted to evaluate the model.

4. Results and discussions

This section provides the relevant results for data analysis of prediction on graduate employment within six months after graduation. Discussions of the results obtained were provided in the form of diagram and tables based on the significant relationship between the duration of employment within six months and learning progress based on eight PLOs using SLR method. We received 47 responses from 216 graduate who completes their study in 2018. The response rate of 22% is acceptable for us to proceed for further analysis.

4.1. Statistical properties of SLR model

Table 4 provides the results of correlation coefficient using Pearson Product Moment Correlation (r), coefficient of determination (R^2), and significant testing (p -value) for the relationship between both duration of graduate to be employed and learning progress.

Table 4. The statistical analysis of employability duration and learning progress

PLO	Correlation, r	Coefficient of determination (R^2)	p -value
PLO 1	-0.0713	0.0051	0.6338
PLO 2	0.0773	0.006	0.6057
PLO 3	0.3197	0.1022	0.0285 ^a
PLO 4	0.0848	0.0072	0.5709
PLO 5	-0.0194	0.0004	0.8971
PLO 6	-0.2250	0.0506	0.1284 ^a
PLO 7	-0.0750	0.0056	0.6165
PLO 8	0.0168	0.0003	0.9109

^a *Significant.

For the relation depicted in table 5, the highest r and R^2 values, as well as the significant values ($p < 0.25$) with $n = 47$, are at PLO 3 and PLO 6. The cut-off value for the p -value is $\alpha = 0.25$ (25%) is based on the mission of National Graduate Employability Blueprint 2012-2017 [17]. Hence, the model generated from both PLO 3 and PLO 6 is more precise to use for prediction purpose than the other models.

4.2. Model validation and evaluation

The model validation results by using error rate analysis indicate a variation of the degree of errors from the dataset that already tested with the model. Based on the error rate of the predicted values, the frequency analysis as shown in figure 5 and figure 6 to evaluate the model performance.

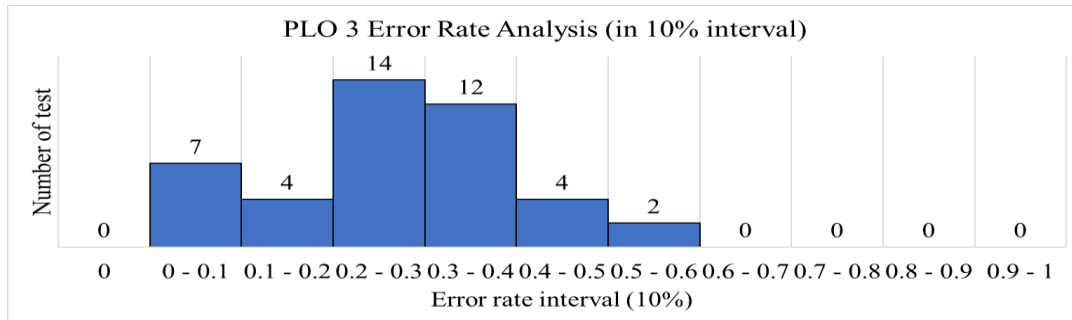


Figure 5. The PLO 3 error rate analysis (in 10% interval)

The distribution of error rate between actual and prediction result in 10% interval range is presented in figure 5. It shows that between the range of 0.2 – 0.3 and 0.3 – 0.4 are the highest error detected in this model. Means that, this model is less correctness in prediction and actual duration.

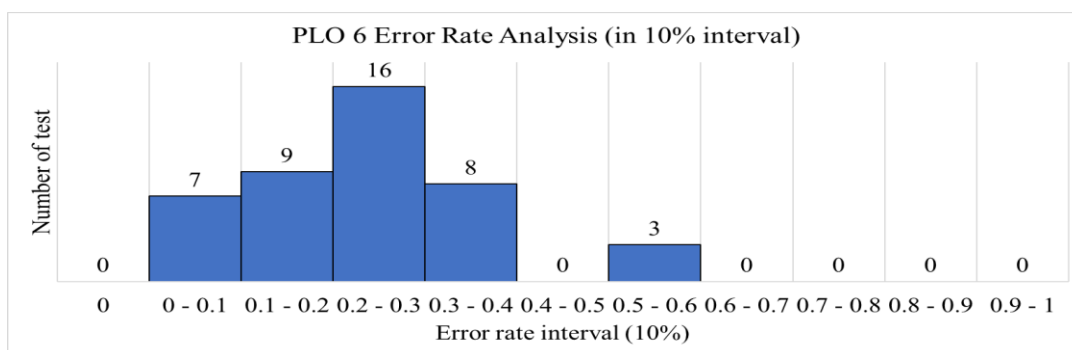


Figure 6. The PLO 6 error rate analysis (in 10% interval)

Figure 6 illustrates the result of PLO 6 model where the error rate is increased at the range of 0.2 – 0.3. This model also shows the same meaning as result in PLO 3. For both models, it analyses the error rate is within 0.6 out of 1%. It can be summarized that this prediction model shows an acceptable or tolerable error rate since it does not exceed 1%. The predicted result as compared to the actual indicates the error is within the targeted timeframe of graduate to get employed.

5. Conclusion

By using simple linear regression (SLR), a predictive model can be developed based on learning progress to predict employability duration to get employed. This method clarifies our first research question (RQ1). As for RQ2, the results show that learning progress of PLO 3 which is '*Social Skills and Responsibilities*' and PLO 6 where '*Problem Solving and Scientific Skills*' shows significant impact on graduate employability performance. The predictive model based on PLO 3 and PLO 6 shows an acceptable error rate that fulfil the mission of National Graduate Employability Blueprint 2012-2017. This suggests that there is high feasibility that the learning progress can be used as predictor for graduate employability performance. Since student learning progress can be obtained from student learning outcomes data, therefore more effective strategies can be plan in improving the quality of curriculum that gives impact on graduate employability performance while they still in study duration.

Acknowledgement

This work was supported by Universiti Malaysia Pahang from grant PGRS1903108.

References

- [1] Shahiri A M, Husain W and Rashid N A 2015 A review on predicting student's performance using data mining techniques *Procedia Comput. Sci.* **72** 414–22
- [2] Parnell A, Jones D, Wesaw A and Brooks D C 2018 *Institutions' use of data and analytics for student success*
- [3] Fitzpatrick J J, Byrne E P and Kennedy D 2009 Making programme learning outcomes explicit for students of process and chemical engineering *Educ. Chem. Eng.* **4** 21–8
- [4] Emily Greenleaf, Megan Burnett P G 2008 *Developing learning outcomes: a guide for University of Toronto faculty*
- [5] Adam S 2004 *A consideration of the nature, role, application and implications for European education of employing 'learning outcomes' at the local, national and international levels*
- [6] Accreditation C for H E 2003 *Statement of mutual responsibilities for student learning outcomes: accreditation, institutions, and programs*
- [7] Wastiau-Schluter P 2004 *Focus on the structure of higher education in Europe 2004/05*
- [8] O'Brien D J 2009 *Writing learning outcomes*
- [9] Kennedy D 2006 *Writing and using learning outcomes: a practical guide*
- [10] H. Taib, S M Salleh, B A Md Zain, M.A. Azlan, S. Mahzan, Hafeez Z A, Ong, P., S. Ahmad, M. N.A Rahman, N.F. Nasir, M. Azham Azmi. H.A Rahman Z N 2017 Programme learning outcomes assessment and continuous quality improvement in Faculty of Mechanical and Manufacturing, UTHM *IOP Conf. Ser. Mater. Sci. Eng.* 165
- [11] Baker R S, HersHKovitz A, Rossi L M, Goldstein A B and Gowda S M 2013 Predicting robust learning with the visual form of the moment-by-moment learning curve *J. Learn. Sci.* **22** 639–66
- [12] Winne P H and Hadwin A F 2013 nStudy: tracing and supporting self-regulated learning in the internet *International Handbook of Metacognition and Learning Technologies. Springer International Handbooks of Education.* pp 293–308
- [13] Molenaar I, Horvers A and Baker R S 2019 What can moment-by-moment learning curves tell about students' self-regulated learning? *Learn. Instr.* 101206
- [14] Yang F and Li F W B 2018 Study on student performance estimation, student progress analysis, and student potential prediction based on data mining *Comput. Educ.* **123** 97–108
- [15] Mahazan A M, Siti Nubailah M Y, Ummi Salwa A B and Wan Mohd Fazrul Azdi W R 2015 Predictors of graduate employability: mediating roles of leadership, ethics, and religiosity *Int. Acad. Res. J. Bus. Technol.* **1** 126–36
- [16] Poornima S and Pushpalatha M 2018 A survey of predictive analytics using big data with data mining *Int. J. Bioinform. Res. Appl.* **14** 269
- [17] Ministry of Higher Education Malaysia 2012 *The national graduate employability blueprint 2012-2017*
- [18] Harden R M 2007 Learning outcomes as a tool to assess progression *Med. Teach.* **29** 678–82
- [19] König R 2009 *Predictive techniques and methods for decision support in situations with poor data quality*
- [20] Pearson K. 1900 On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling *Philosophical Mag.* **50** 157–75