IOP Conference Series:
Materials Science and
Engineering

PURPOSE-LED
PUBLISHING™

PAPER • OPEN ACCESS

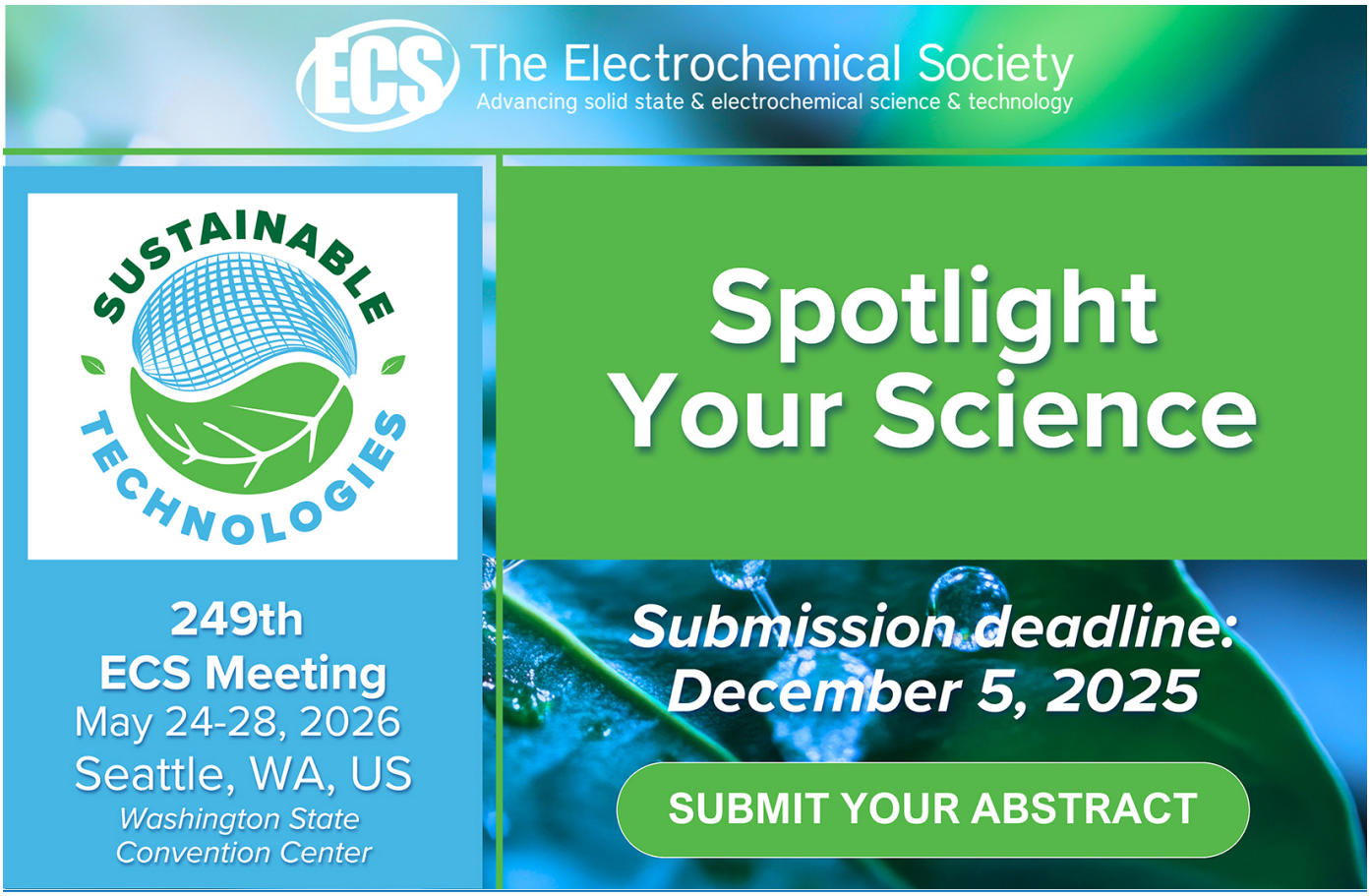
Predicting Graduate Employability based on Program Learning Outcomes

To cite this article: Wan Nor Afiqah Wan Othman *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **769** 012018

View the article online for updates and enhancements.

You may also like

- Student Learning Progress as Predictor for Graduate Employability Performance
Wan Nor Afiqah Wan Othman and Aziman Abdullah
- Use of Electrochemical ByProducts in Glass Making: I . Arsenic, Selenium, Tellurium
Alexander Silverman
- A comparative Study on Graduates' Employment in Malaysia by using Data Mining
Nur Iman Natasha Binti A'rifian, Nur Sakinah Amirah Binti Mohd Daud, Athirah Faiz Binti Muhamad Romzi et al.



This content was downloaded from IP address 190.57.151.78 on 05/11/2025 at 12:38

IOP Conf. Series: Materials Science and Engineering **769** (2020) 012018 doi:10.1088/1757-899X/769/1/012018

Predicting Graduate Employability based on Program Learning Outcomes

Wan Nor Afiqah Wan Othman, Aziman Abdullah, and Awanis Romli

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. Many studies based on the literature and adopted approach by Ministry of Higher Education regarding graduate employability are using survey. This approach is lack with on-demand analytical capability for impactful decision making. There is a lack of study that predicts the duration of graduate to get employed based on quantitative analysis. Since all institutions of higher education are compulsory to adopt and implement outcomes-based education (OBE), this study aims to develop a predictive model on GE based on program learning outcomes (PLO) data. There are two data sources used in this study, institutional academic database and online feedback from graduate. This study used simple linear regression to measure the degree of relationship between the category of PLO with the duration of graduate to get employed. This study received 47 responses from 216 with a response rate of 22%. PLO1 and PLO6 which are '*knowledge*' and '*problem solving and scientific skills*' respectively show high significance values on the duration of graduate to get employed. The linear models developed based on PLO1 and PLO6 were validated with error rate analysis and evaluated with error rate frequency analysis. The results show the model has potential value to be used to predict graduate employability performance within the time frame (6 months) as determined by Ministry of Higher Education. With prediction capacity from the developed model, more intervention program can be strategically planned to assure graduate can be employed in time and in-field.

1. Introduction

Graduates' employability is issues that have been rise to many policy implications for higher education in many developing countries such as Germany, United Kingdom, China, Vietnam as well as Malaysia [1–9] especially in securing in-field employment [10]. Several factors are contributing to the graduates' employability issue [3]. The main is an individual factor regarding to the skills gap, between what has acquired and what is been expected. It refers to the graduate's talent aspect in promoting their skills not only knowledge-discipline skills but also include soft-skills as well. This issue becomes more critical where World Economic Forum claims that "demand for various skills in 2020, more than one third (36%) of all jobs across all industries require complex problem-solving as one of the core skills, rather than 1 in 20 jobs (4%) for physical abilities" [11]. The employability includes all the aspects of individuals proactive behavior, personal management attitude and career management aptitude in order to get a necessary job [12].

Tracer study is a current approach as an assessment tool to collect and trace graduate's data through standardize survey. This approach aims are to gain experimental and observational data that give the impacts to the education programme and improve the transition of graduates from education to the labour market [13–18]. In Malaysia, Ministry of Education through Department of Higher Learning

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

has conducted a yearly National Graduate Tracer Study. It is a web-based survey system collaborate with all higher education institutions (HEIs) upon graduation to improve the graduates' employability since 2006 [19]. Tracer study should be a tool that benefits stakeholders. However, there are several drawbacks which involved with key problems of data quality and lack of insight to the right audience. This data is also limited and restricted the institutions to provide practical improvements on the impact of universities' courses to the related field in industries [17], [18], [20], [21]. Hence, the lack in real time of decision making, comprehensive data and inclusive graduate data respectively of tracer study should be emphasized.

This study aims to develop a predictive model for graduates' employability duration to get employed using Simple Linear Regression (SLR) method. The key research questions are stated as below:

RQ1: Which PLO's that highly influence the duration of graduate to get employed?

RQ2: Is it feasible to predict the duration of graduate employability?

2. Literature review

2.1. Malaysia Qualification Framework (MQF)

The Malaysian Qualifications Framework (MQF) which was approved under the Malaysian Qualifications Agency (MQA) Act, 2007, is the instrument which sets national classifications of qualifications, levels of learning achievements based on learning outcomes, prescribed academic load at each level and is associated with the title of a named qualification. All academic programme in Malaysia higher education will be accredited only when it is compliance with the framework. The framework defines there are at least eight standard program learning outcomes (PLO) that any program must comply with. Table 1 shows the attributes of eight domains in PLO

Table 1. The eight standard 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

The purpose of the framework is to assure the quality of higher education. One of the quality issues in higher education is graduate employability. It is expected to offer much better insight for various stakeholders as been mentioned by Ministry of Higher Education (MoHE) in Malaysia Education Blueprint 2015-2025 to address the issue on graduates' employability.

2.2. Prediction on Graduate Employability

Zwick & Sklar [22] applied a linear regression model to predict first year college grades (FGPA) as well as degree completion using previous school grades and SAT scores. They concluded that the result in evaluating the contributions of SAT score and GPA is statistically significant in the prediction of graduation. Another study [23] also adopted linear regression analysis to investigate the predictors of graduates' employability based on eight graduates' skills. From the hierarchical regression results, it shows the eight predictors explained 37% variance of employment status. The study concludes that value-related graduates' skills are vital for employability. Based on the related works, this study also adopts simple linear regression method in predicting graduate employability based on the program learning outcomes data.

3. Methodology

Figure 1 shows the research model adopted in this study. Predictive Analytics Process adapted from the actual version (refer [24]) was used.

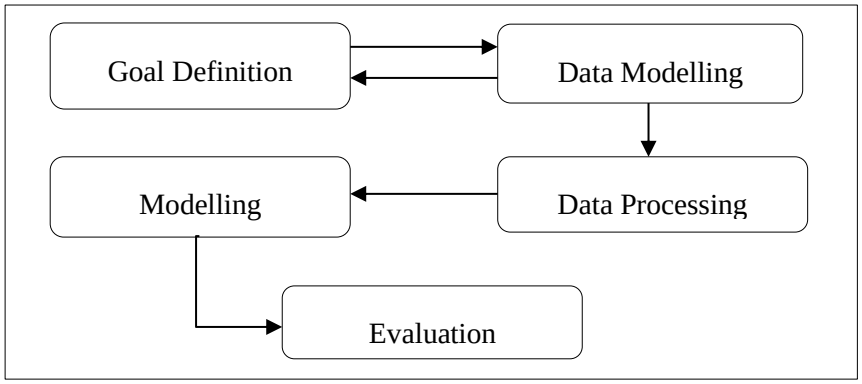


Figure 1. Predictive analytic process – modified version from Poornima & Pushpalatha (2018)

3.1. Phase 1: Goal definition

The goal defined for this study is to identify which competency 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.

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 graduate. Table 2 illustrate the 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 graduation.	Online feedback.

3.3. Phase 3: Data processing

Time duration is a data value that dynamically changes over time itself. We only asked the graduate to response simply the first date that they are get employed for the first time after graduation. 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).

$$Duration = Date\ of\ First\ Job - Date\ of\ Senate\ (graduation) \quad (1)$$

Table 3 shows the structure of the data how duration of employed graduate is been produced. This study adopts a common spreadsheet application with 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 Senate	Date of First Job	Duration (Number of Days)
Student 1	2018-07-25	2018-08-01	7
Student 2	2018-07-25	2018-01-07	166

Lastly, data normalization process involved where it classifies values of data by using minimum and maximum steps. For example, within scales 0.0 to 1.0.

3.4. Phase 4: Modelling

Modelling phase can be done by training and testing data into statistical method or algorithms to create the model. The model of candidates, selection and validation are included in this phase. In this study, we select simple linear regression (SLR) as a method that will be used to build the model. The reason why SLR been used is because it is simple, direct, effective and easy to understand [25].

Model of candidates is a process to derives 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 (eight PLO) of each candidate. Pearson Product Moment Correlation (r) [26] is used. The significance testing (p-value approach) is used to make conclusions of hypothesis where p-value are compared to a significance level (α). To summarize:

p-value < α => reject H_0 => accept H_a

p-value ≥ α => fail to reject H_0

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

Model selection process produces the best SLR model to be choose among the eight PLOs. In order to get the best model to use for the testing and forecasts of upcoming data sets, the statistical method need to 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.

Model validation process is implemented when the selected model is tested using same dataset to identify the prediction accuracy performance. It is to prove the attributes used in this study able to build the correct models. The accuracy performance of model is validated based on the error rate of the predicted values. The error rate of the model is defined as:

$$\textit{Error Rate (\%)} = (|\ \textit{Predict-Actual}\ |\ /\ \textit{6 Months}) \times 100\% \quad (2)$$

where,

Predict = Prediction duration using SLR model

Actual = Actual duration from raw data

6 Months = Approximately 182 days

3.5. Phase 5: Evaluation

The evaluation phase is to measure the effectiveness of model in UMP as a case study. The model was evaluated with frequency analysis of error rate.

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 eight PLO using SLR method.

4.1. Statistical properties of SLR model

Table 4 provides the results of correlation coefficient for the relationship between both variables using Pearson Product Moment Correlation (*r*). This *r* value aims to determine the strength of relationships among both variables either positive or negative correlation.

As shown in table 4, the correlation result of duration against PLO 7 in Pre-LI group achieved the highest *r* values at 0.2587 compared to others. There are weak negative and positive correlation was analyzed. In this study, the positive correlation referred to the result of duration and PLOs variables are both increased. It means students took a long time to get a job after graduation even though they got high-grade points for their PLO. This happens due to the dataset of which almost 47 candidates get high-grade points or there are some zero values in Pre-LI and LI groups of each PLOs. In this research, the expected relationship is a negative correlation where better PLO attainments, the less time for fresh graduate to secure a job offer.

Table 4. Pearson product moment correlation of the relationship between duration and PLO 1 to 8 for Pre-LI and LI

PLO	Pre-LI	LI
PLO 1	-0.2144	-0.2027
PLO 2	-0.0776	0.0524
PLO 3	-0.1698	0.0718
PLO 4	-0.1465	-0.0107
PLO 5	-0.0858	0.04941
PLO 6	-0.2515	-0.2022
PLO 7	**0.2587**	-[a]
PLO 8	-0.0700	-0.0669

[a] '-' No data were found for each student in PLO 7 (LI group).

The results of coefficient of determination (R^2) is provided in table 5. The highest R^2 value is also on PLO 7. The R^2 = 0.0669 means that around 7% variation in duration can be explained by PLO 7 in Pre-LI group. This means that the model generated from the relationship between these two variables is more precise to predict than the other models.

Table 5. Coefficient of determination of the relationship between duration and PLO 1 to 8 for Pre-LI and LI groups

PLO	Pre-LI	LI
PLO 1	0.046	0.0411
PLO 2	0.006	0.0027
PLO 3	0.0288	0.0052
PLO 4	0.0215	0.0001
PLO 5	0.0074	0.0024
PLO 6	0.0632	0.0409
PLO 7	**0.0669**	-[b]
PLO 8	0.0049	0.0045

[b] '-' No data were found for each student in PLO 7 (LI group).

The significance testing by using *p*-value is implemented to check the significance of the independent or predictors in the model. Table 6 shows the significant values of both variables with a

cut-off value for the *p*-value is α = 0.25 (25%). This case is based on the mission of National Graduate Employability Blueprint 2012-2017 [27]. The null hypothesis, in this case, is H_0: $\beta_1 = 0$, i.e., that there is no significant linear relationship between the duration and PLO 1 to 8 for Pre-LI and LI groups.

Table 6. Significant testing of relationship between duration and PLO 1 to 8 for Pre-LI and LI groups

PLO	Pre-LI	LI
PLO 1	**0.1479**[*c]	**0.1717**[*c]
PLO 2	0.6043	0.7265
PLO 3	0.2539	0.6313
PLO 4	0.3259	0.9433
PLO 5	0.5662	0.7415
PLO 6	**0.0882**[*c]	**0.1730**[*c]
PLO 7	**0.0791**[*c]	-[d]
PLO 8	0.6403	0.6551

[c] * Significant.
[d] '-' No data were found for each student in PLO 7 (LI group).

As shown in table 6, the significant values ($p < 0.25$) with n = 47 for the relationship between duration and PLO 1 to 8 refers to the bold values. However, there is insufficient evidence to conclude that PLO 2 until 5 and PLO 8 for both groups are significant linear relationship to the duration. It because of the correlation coefficient is not significantly different from zero. Therefore, do not reject the null hypothesis, H_0: $\beta_1 = 0$.

4.2. Model validation and evaluation

The results from model validation with error rate analysis show variation of the degree of errors from the dataset that been tested with the model. Based on the error rate, we evaluate the performance of the model by performing frequency analysis using a histogram chart as shown in figure 2 and figure 3.

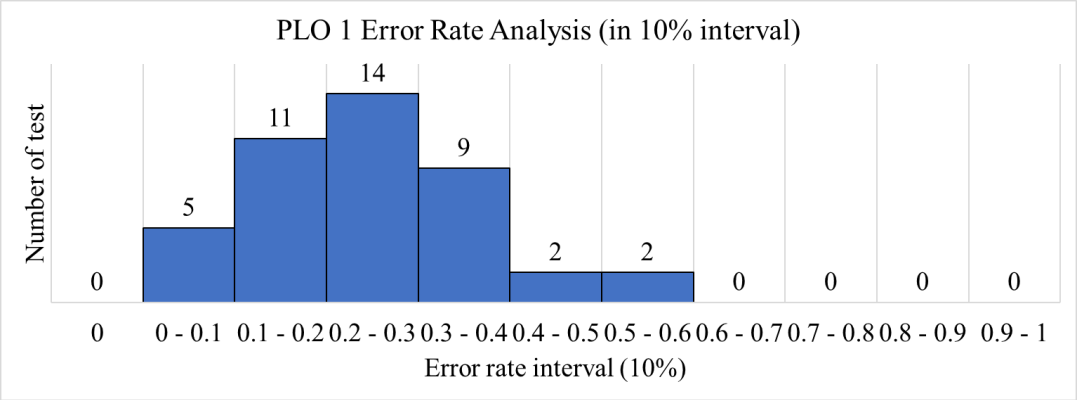


Figure 2. The PLO 1 error rate analysis (in 10% interval)

Figure 2 shows the distribution of error rate between actual and prediction result in 10% interval range. As the error rate of PLO 1 shown in figure 2 increases at the range of 0.2 - 0.3, this model minimally predicts the correctness of the actual duration.

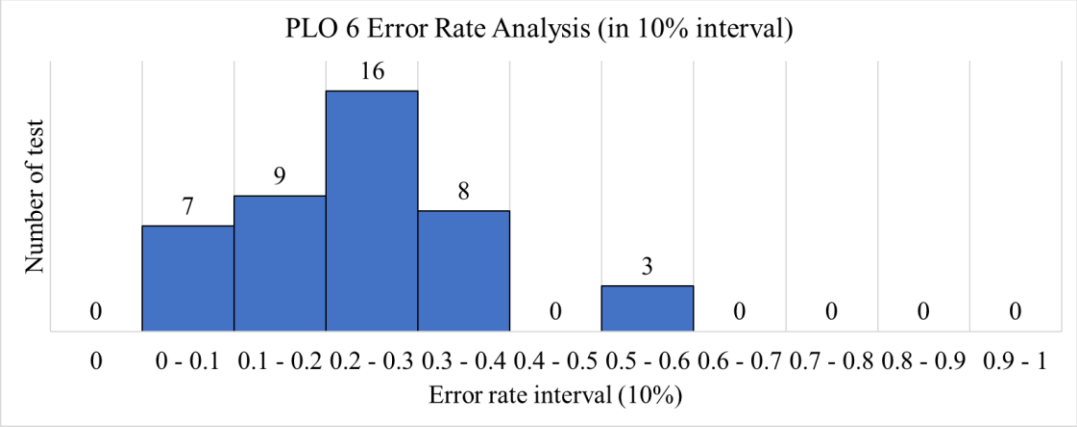


Figure 3. The PLO 6 error rate analysis (in 10% interval)

Figure 3 shows the PLO 6 pattern of error rate analysis is approximately the same as PLO 1. For both analyses, it found that up to 0.6 out of 1% have the errors. In this case study, it considered as better forecasting models where the error rate does not exceed 1%. This also means it accurately predicts the duration to get employed before six months after graduation.

5. Conclusion

As a conclusion, there is high feasibility that the duration of graduate gets employed after finishing their study can be predicted based on the learning outcomes attainment data. According to the findings, it can be summarized that PLO1 and PLO6 show highly significant value on the duration of the graduate to get employed. The negative correlation between both PLO's against duration to get employed indicates that the more competent graduate on those skills, the more likely they can get employed earlier. Furthermore, the result from model validation and evaluation support this conclusion.

By identifying this competency that highly affect graduate employability, more student intervention and enrichment program can be done for student before they go for an industrial internship. Hence, it ensures that students are more ready for the internship compares to before. This model, however, may only applicable for the academic program that has an industrial internship at the last semester of study.

For future research, this study is required to develop comprehensive models able to predict the graduates' employability for all academic programs. Besides, a small dataset was found which need to increase the number of respondents in order to get high response rate. Future work also should carry out the deployment phase. This phase is a generating report of the outcomes gets through the models. It produced for university reference as this is the valuable information that can give a better decision making.

6. Acknowledgement

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

References

[1] Department of Statistics M 2018 *Labour force in Malaysia*
[2] Bai L 2015 Graduate unemployment : dilemmas and challenges in China's move to mass higher education *China Q.* **185** 128–44
[3] Yusof N and Jamaluddin Z 2015 Graduate employability and preparedness : a case study of University of Malaysia Perlis (UNIMAP), Malaysia *Geogr. J. Soc. Sp. 11* 129–43
[4] Sumanasiri E G T, Yajid M S A and Khatibi A 2015 Review of literature on graduate employability *J. Stud. Educ.* **5** 75
[5] Tran T T 2015 Is graduate employability the 'whole-of-higher-education-issue'? *J. Educ.*

Work **28** 207–27
[6] Zheng J, Tang J and Gao J 2012 Construction of graduate employment service system based on public information platform *AASRI Procedia* **1** 410–5
[7] Leo M 2016 *What you didn't know about fresh graduate unemployment in Malaysia [infographic]*
[8] ILO 2013 *Global employment trends for youth 2013* vol 1
[9] Shao H R 2011 *A study of the employability-based curriculum development in american universities : focus on business schools*
[10] Futch K W 2014 *Factors that predict time to in-field employment of associate degree graduates: a study of one college in the technical college system of georgia*
[11] World Economic Forum 2016 The future of jobs employment, skills and workforce strategy for the fourth industrial revolution *Growth Strateg.* 2–3
[12] Clarke M 2008 Understanding and managing employability in changing career contexts *J. Eur. Ind. Train.* **32** 258–84
[13] Aina L O and Moahi K 1999 Tracer study of the Botswana library school graduates *Educ. Inf.* **17** 215–44
[14] Kyaw U K 1992 *A tracer study of recent graduates : implications for education and manpower planning*
[15] Manda M and Matidza I 2016 A tracer study of 2011-2015 graduates from the land management programme at Mzuzu University *J. Educ. Soc. South. Africa* **3** 45–67
[16] Damotharan L V 2012 *Tracer study on AIMST University students using data mining*
[17] European Training Foundation 2017 *Tracer studies evaluating the impact of training programmes*
[18] Schomburg H 2016 *Carrying out tracer studies guide to anticipating and matching skills and jobs* vol 6
[19] Latif L A and Bahroom R 2010 OUM's tracer study: a testimony to a quality open and distance education *ASEAN J. Open Distance Learn.* **2** 35–47
[20] Rogan M and (NALSU) N A L S U 2016 *Tracing graduates into the labour market*
[21] Sanchez M P R and Diamante V A M 2017 Graduate tracer study of the college of nursing *Malaysian J. Nurs.* **8** 41–7
[22] Zwick R and Sklar J G 2005 Predicting college grades and degree completion using high school grades and SAT scores: the role of student ethnicity and first language *Am. Educ. Res. J.* **42** 439–64
[23] 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
[24] Poornima S and Pushpalatha M 2018 A survey of predictive analytics using big data with data mining *Int. J. Bioinform. Res. Appl.* **14** 269
[25] König R 2009 *Predictive techniques and methods for decision support in situations with poor data quality*
[26] Pearson K. 1900 On the criterion that a given system of deviations from the probable in the case of a correlated system of varibales is such that it can be reasonablly supposed to have arisen from random sampling *Phylosophical Mag.* **50** 157–75
[27] Ministry of Higher Education Malaysia 2012 *The national graduate employability blueprint 2012-2017*