Sub-theme: Inspiring Innovation

Title: A Study on Learner Retention and Academic Performance using Student Data and Technologies Towards Building a Resilient Open Education System and Innovative Solutions

Thirumeni T Subramaniam*,
thirumeni@oum.edu.my,
Open University Malaysia

Nur Amalina Diyana Suhaimi,
nad@inscale.net,
INSCALE Asia

Rohaizak Omar,
rohaizak@oum.edu.my,
Open University Malaysia

Abstract

This study on learner retention and academic performance is designed to utilize the rich learners' database towards improving the education services offered by the Open University Malaysia. The objectives of this research are: (i) to identify possible trends in learner retention and academic performance; (ii) to identify possible factors that could influence learner retention and academic performance; and (iii) to develop an infographic system. The research requires mining of learners' data and data analysis. During the first phase of data mining and data integration, a number of issues had to be resolved by involving different departments at the university. Data identification, data cleaning and coding was done to enable the achievement of the targeted objectives. The move to the use of 'big data' promotes transformation of institutional research projects at the university. Research methods employed also varies from previous survey method to graphical analysis, explorations, and improvement of data structures. Identification of students at risks in terms of retention and academic performance enables the university to make evidence-based decision making and provide targeted solutions. While, the transformation in institutional research offers numerous opportunities, there are also numerous challenges. Findings are presented using selected evidences with attention to the transformation in the research method and the benefits offered by such transition. The proposed transformation, and possible innovations of the mechanisms through which quality education services can be offered are discussed. Depicted findings along with the transformation that are in place could facilitate efficient evidence-based decision-making processes at the university. Use of machine learning and analytics software in research are being explored at present to develop auto-detection of the impact of an introduced solution, predictive modelling (retention, and academic performance) and real-time data visualization. Concurrent research efforts are also in place to promote innovations in: i) pedagogical processes, (ii) learning environment (process), (iii) education learning materials (product), and (iv) administration processes supporting the learners.

Keywords: Learner Retention, Academic Performance, Research Transformation, Use of Technology, Research Methods

1.0 INTRODUCTION

Open and Distance Learning (ODL) institutions have always thrived to be in the forefront in the efforts providing efficient and effective management of education delivery. While technology-enhanced education is always considered an important aspect of education delivery in an ODL setting, various ways technology is used and can be used in education delivery varies among the ODL institutions. An example is the study conducted by Goitsemang and Seretse (2019). Effective use of available technology is influenced largely by the socio-economic factor within a locality (Shavkun, Bukharina, Dybchynska, and Onyshcchenko, 2021). Nevertheless, the diversity in the way technology is used in managing education delivery opens windows for creativity and innovations. Innovations in education can be studied through research in the following areas: (i) pedagogical process, (ii) learning environment (process), (iii) education learning materials (product), and (iv) administration processes supporting the learners. Research in these aspects ideally should be guided by the 4th Sustainability Development Goal (SDG4) to provide Quality Education.

Open University Malaysia (OUM) have been involved in research in ODL in the stated four areas. In this paper, we would explore the research in the fourth area, administration processes supporting the learners. In this aspect, a focus area for many years has been in student retention where research was focused on the use survey method to study factors influencing learner retention. In 2019, the university widen its research interest to other disciplines and thus opened the window for transdisciplinary research. This move changed the research methodology used in the study of learner retention from survey to the use of ‘big data’. At the onset, only the dataset fell within the ‘big data’ scope, the methodology remained traditional via use of traditional analysis and use of statistical tools. Today, work has been initiated in the area of programming, machine learning, modelling and infographics. In this paper, both the findings from the study and the transformation of institutional research to explore data analytics will be discussed.

Research in retention using the ‘big data’ enables the use of existing data to understand trends about OUM learners: enrolment, attrition, progress through the semesters, academic performance, and graduation. Attrition has always been a critical issue for many ODL institutions where the learners are mostly working adults. It would be beneficial for ODL institutions to be able to predict learners who are most likely to drop out so as to take the necessary remedial actions to prevent it from happening. There are two main factors that can contribute to attrition, namely the demographic and the entry qualification are found from learners’ database. The demographic factor includes gender, age, occupation, income and so on. Whereas, the entry qualification can be analysed from the factors relating to competency in language and mathematics skills that can influence learning (literacies). Another possibility that can be explored is the learners’ academic performance in completing compulsory courses at the early semesters and how it influences their retention. The trends found can be used to highlight critical points concerning learner’s difficulty in persisting with their studies. Such findings are important in designing an effective learner support system that can enable learners to persist and complete their programme within a reasonable time-period. This is crucial in building a resilient education system. Correlating such trends with demographic data as well as entry qualifications can indicate the factors that enable learners to persist in their studies. Such findings can assist the university to use the factors to predict the likelihood a learner may require additional support in completing their studies.

2.0 LITERATURE REVIEW

Open and distance education, which at present is almost synonym with online education has become increasingly popular. The relatively low retention rate among online open and distance institutions suggests the importance of retention efforts in the field of enrolment management. The importance of learner retention efforts in enrolment management was also highlighted by Perry and Sullivan (2007). They raised the fact that learner retention is also recognised as a barometer of learner satisfaction with life and learning. Thus, any effort to understand the dynamics between learner enrolment, learner retention, and academic performance (Cumulative Grade Point Average (CGPA) and on-time completion rate, and graduation rate) is crucial in identifying both strength and weaknesses in the academic experience offered (including the learning environment within which an academic programme is delivered) by the learning institution.

One crucial factor to be considered is the learners themselves. A learner, influenced by a number of different dimensions is a highly important factor in determining his or her potential for academic success. ODL institutions often design their education offered based on Andragogy which is based five assumptions about adult learners: (i) autonomous, (ii) benefit from experiential learning, (iii) high learning readiness, (iv) problem-centred in learning, and (v) learning motivations are derived from internal factors. The criticism that is well presented by Cercone (2008) argues the validity of such assumptions. Thus, suggesting the need to create a more supportive learning environment. The needs of each learner can be understood by exploring available secondary data of the aforementioned demography and entry qualification factors. Data analysis can reveal the vital relationships between enrolment, retention, graduation and academic performance. This dynamic relationship is known to be affected by several important factors that require a thorough evaluation. Tresman (2002) proposed a theoretical framework that is based on several models including the Quality Improvement Model and Tinto's Model. Learning about learners' profile and linking them to the available secondary data could expand this study, resulting in a more comprehensive research related to learner management. This phase is based on action analytics which can be used to improve learner retention and success. According to Norris, Baer and Offerman (2009) such action analytics can be carried out through a number of different mechanisms: longitudinal data analysis, predictive modelling and dynamic analysis. In addition, the analytics will enable effective decision making in reviewing the performance of the offered programmes.

The background information about learners is useful, but not comprehensive in enabling the institution to predict retention of a learner. There is a need to understand the reasons behind the trends that are observed. A number of studies have been carried out to understand why some learners face barriers in their learning. However, the barriers for any learner population given the difference in culture, background and other influences can be different. As a result, there is a need to carry out qualitative studies to identify the underlying factors that could influence the ability of a learner to retain in the system and attain promising results in their academic performance. Tung (2012) emphasized the need to have a clear understanding of factors that contribute to learner attrition before any intervention can be formulated. Predicting learner retention in such setting, require greater attention to the method of analysis. The importance of digitisation of existing operations particularly under the New Normal was highlighted in the July 2020 playbook by the World Economic Forum. This drives the motivations to explore the use 'big data' and possibly the technologies associated with data science.

Efficient data monitoring could help the institution to design and implement Proactive Interventions. Much of the needed inventions is related to teaching and learning improvements. Crosling, Heagney and Thomas (2009) have highlighted some of the teaching and learning improvements that could possibly improve learner retention in higher education. One initiative that is being explored at OUM is the virtual orientation programme. A second intervention that is being proposed at the institution is a programme in a number of literacies that will not only help to improve learner retention, but also to improve 21st century literacy skills among the public. This initiative is also expected to improve learner's academic performance. The positive influence of literacy in improving both retention and academic performance is highlighted in a study by Glew and Salamonson (2019).

The objectives of this research are:

- (i) to identify possible trends in learner retention and academic performance;
- (ii) to identify possible factors that could influence learner retention and academic performance; and
- (iii) to develop an infographic system.

3.0 METHODOLOGY

This research focuses the use of analytics, and is divided into two phases. The first phase is the Provision of Data Phase, and the second is the Interpretation and Visualisation Phase (Powell and MacNeill, 2012).

<u>First Phase (Data Management)</u>

The first step in the first phase of this study requires data extraction/retrieval that was carried out with assistance of a programmer from OUM's Group Information, Communication and Technology Support (GICTS). Data from two systems (Learner Records and Assessment) were integrated. The systems were managed by both the Admission and Record Unit (ARU), and the Assessment and Examination Department (AED). Extracted dataset was reviewed and studied with the help of a staff from ARU.

Next, the dataset was cleaned in consultation with several units (GICTS, ARU and AED).

Thirdly, the cleaned dataset was transformed for use in the next stage.

Ideally, further data integration should occur at this stage. A learner profiling system is being developed by another research group will support this project at a later phase.

Second Phase (Data Analysis and Data Presentation)

The second phase was carried out using Microsoft Excel and the statistical tool, SPSS version 22. The analysis led to the discovery of:

(i) trends/patterns in learner enrolment, retention, deferment, registration status, academic performance and graduation;
(ii) possible factors that could identify learners who are not persistent in their study and those at risk of attrition due to their CGPA values; and
(iii) depiction of quality of programmes using infographics.

This project is being developed to enable a time series analysis. Findings from this research will serve as an important output in understanding the trend in both learner enrolment, progress and performance.

The population for this study consists of all OUM learners from the first intake in September 2001 to January 2018.

4.0 DATA ANALYSIS

Before the findings are presented, several challenges that are important to the work are presented here.

- Data extraction is not automated and is both tedious and time consuming. At present, machine learning is being explored as a solution to perform data mining, a process to extract useable data from a large set of raw data.
- Data cleaning is only done partially. The university is taking a collective and concentrated effort to populate the student data, detect (autodetection) and prevent possible errors.
- Data transformation is also not automated at present. This will be done using machine learning.
- Data integration from different system were explored. It will be performed once the profiling system is ready. Additional dimensions will be explored and measures to integrate data will be carried out if the identified factors are found to influence learner retention and/or academic performance.
- Data exploration and analysis were limited to the use graphs from excel files and correlation data from SPSS. At present, the use of analytics software such as Tableau and Rstudio is being explored.
- The use of data analysis software functions to model and predict learner retention and academic performance is being explored.
- The development of infographics was also not automated. The use of data analysis software functions for data visualisation is being explored.

Findings from the data exploration and analysis are presented here. Data were analysed by programme levels: Diploma, Bachelor, Postgraduate Diploma, Masters and PhD. However, for the purpose of illustration only trends from the Postgraduate Diploma in Teaching (PGDT) with a total number student of 751 are depicted in this paper. PGDT is the only programme at the Postgraduate Diploma level. The programme was first offered in January 2011.

4.1 Trends

A number of trends were depicted by analysing the data. Useful trends that were analysed range from:

#Intake Trend (Figure 1). As expected, a programme takes few years (at least three) to mature.

#Registration Profile of an intake (September 2015) across the semesters (highlighting how the students progress through a programme) (Figure 2). This trend remains the same regardless of the intake. One of the patterns identified from the analysis is that the number students defer remains highest during the third semester for any intake and any programme. The profile of PhD programme (possibly all research programme) differs from the observed trend.

#Trends of identified factors such as the demographic factor identifying the location of learners through the learning centres. Ideally, each factor could also be depicted by a particular intake, and variation across time can be depicted using the right infographics. Understanding of such variation over time would help to differentiate the influence of the analysed factor on the tested dependent variables: learner attrition and academic performance. (Figure 3)

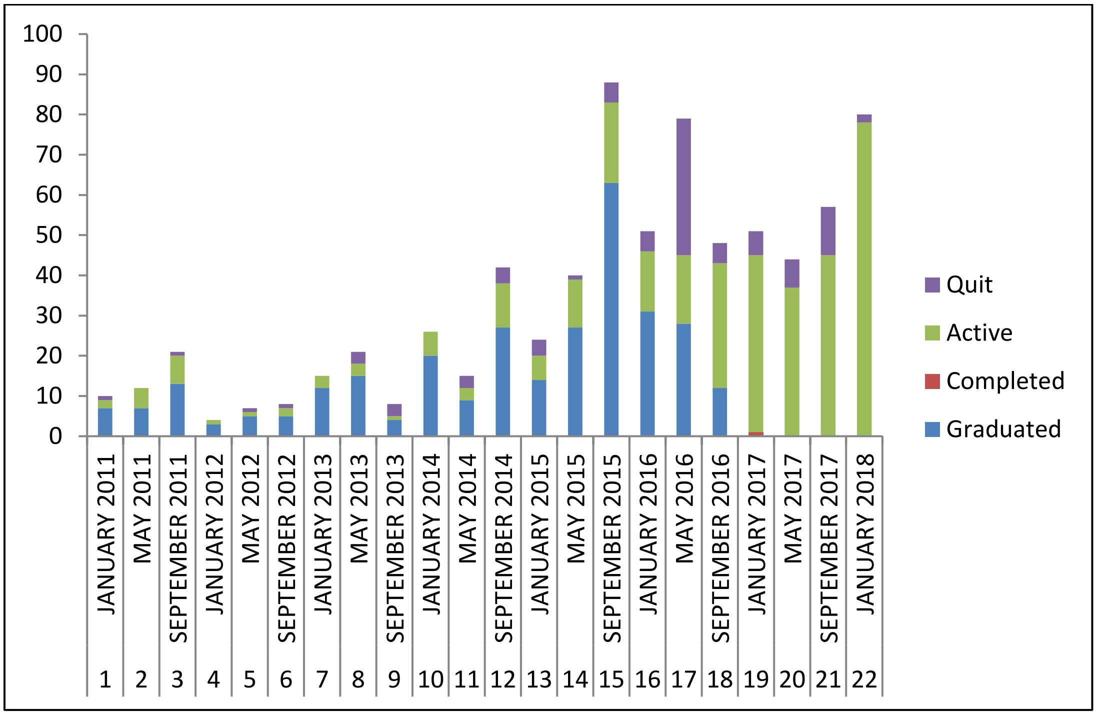

Figure 1: Intake Trends with details of learners who have graduated, who are active (includes those who have completed but have yet to graduated) and those who have quit the programme

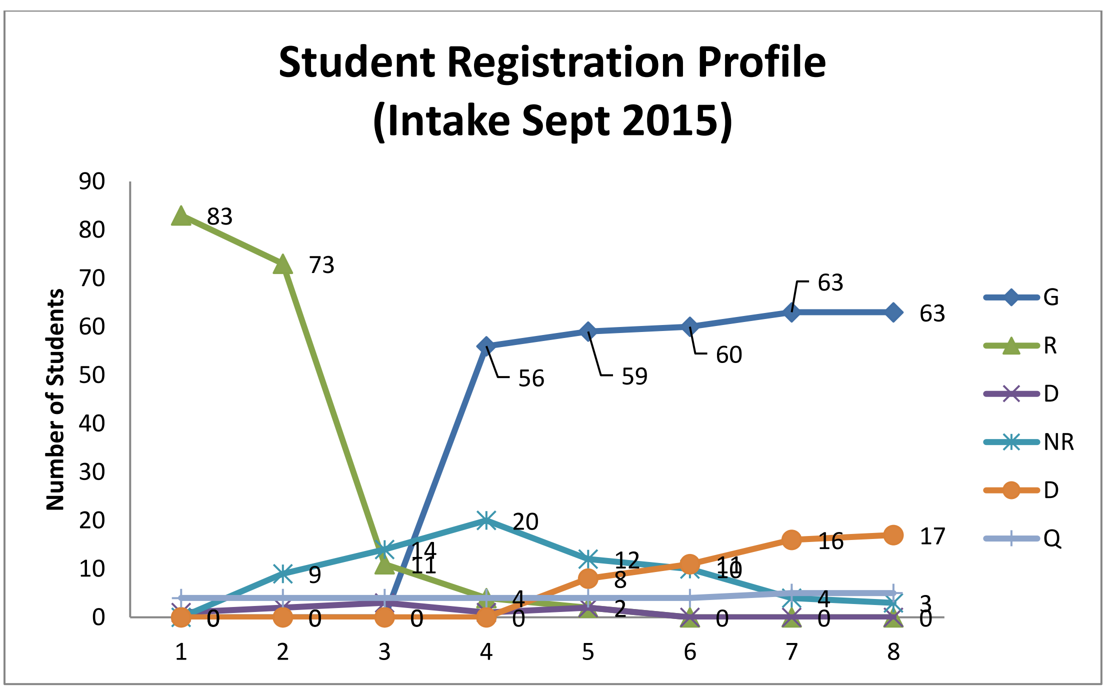

Figure 2: Registration Profile (Graduated, Active (Registered, Defer, Not-registered or Dormant), and Quit)

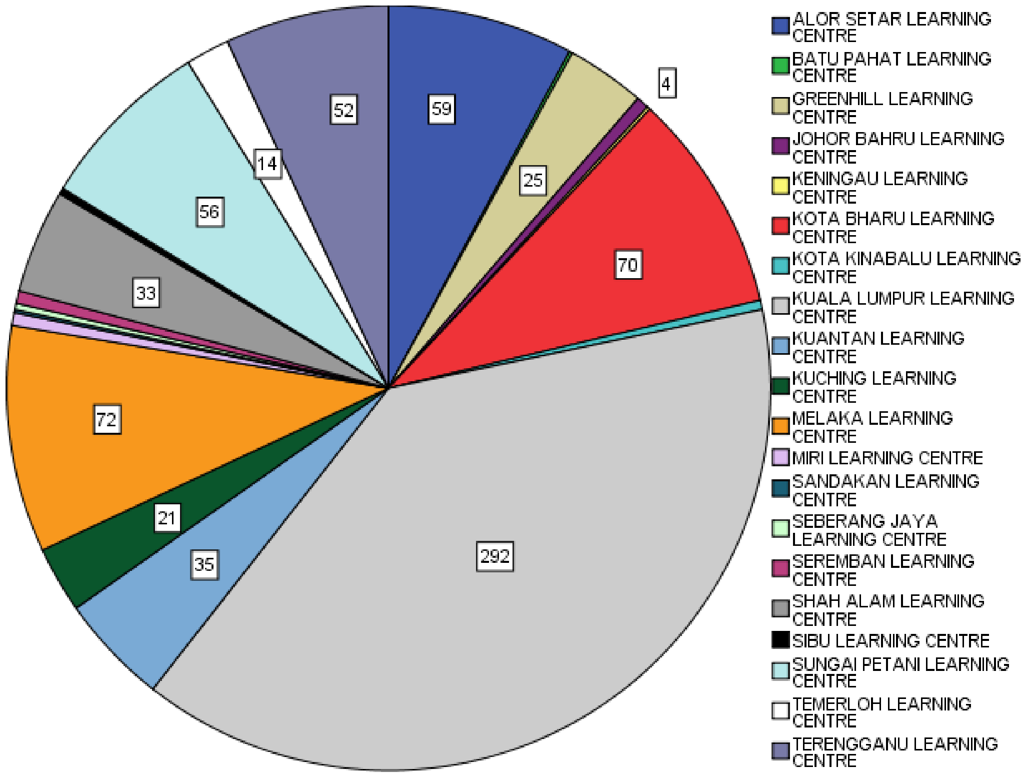

Figure 3: Distribution of PGDT student across the Learning Centres

4.2 *Analysis of Dependent Variable and Its Correlation with Independent Variables (Factors)*

In the case of learner retention, the values indicate excellent performance in the PGDT progamme at 93.6%. The average retention rate for programmes across various levels is around 90% with exception of the PhD programmes. However, a closer look at how the various profiles under the Active status profiles (Completed, Register, Defer, Not Registered and Dormant) changes, shows that the actual rate might be much lesser at 71.6% assuming none of the dormant learners (learners who do register for more than 3 semesters) graduate. The university do take measures to reach out to the dormant student, and there few learners have returned to the programme and graduate. An excellent feature of the PGDT programme is that 63.6% of its total graduates have graduated on time; which is a rather high percentage (88.8% of the total leaners who have graduated (71.6%)).

Table 1: Table for depicting the Performance of the PGDT Programme

	1	2	3	4	5	6	7	8	Total	%	
Graduated	0	0	0	56	59	60	63	63	63	63.6% 71.6%	G G*
Completed	0	0	56	3	1	3	0	0			
Register	83	73	11	4	2	0	0	0			
Defer	1	2	3	1	2	0	0	0			
Not Registered	0	9	14	20	12	10	4	3			
Dormant	0	0	0	0	8	11	16	17	20	22.7%	NR/D
Quit	4	4	4	4	4	4	5	5	5	5.7%	Q
Total	88	88	88	88	88	88	88	88	88	100%	

Note: G stands for On-time-graduation, and there were learners who graduated after 8 years being in the programme.*

Whereas in the case of academic performance, the tracking of the final CGPA was tedious due to the way the data was extracted. The mean for the September 2015 intake was only at 3.36, however the distribution between the students with CGPA below 3.5 and above 3.5 was even. The tabulation of data for CGPA is normal across the programmes.

However, the correlation test between both the dependent variable to the factors studied revealed no correlations. An example of such test using Pearson's Correlation between CGPA in the fourth semester and Grade Point Average (GPA) values in the first three semesters (GPA 1, 2, and 3 respectively) shown in Table 2. The limitation of the test is due to high number of missing data (demographic data), and uneven sampling. The challenges in predicting learner retention even in the case where data mining was used was highlighted by Bilquise, Abdallah, and Kobbaey (2020). Another weakness of this study is limitation of the number of dimensions explored (only two). However, the opportunities offered by the research is worth the explorations. The 2018 report by Lougheed, Drinkwater, and Jamieson is a good read in understanding the opportunities and challenges in developing a predictive model for learner retention.

Table 2: Table for depicting the Performance of the PGDT Programme

			n	**Remarks**
INTAKE JAN 2011 – SEP 2011				
2.4.1	**GPA 1 -> CGPA 4**	Pearson's R = 0.122	43	No correlation
	GPA 2 -> CGPA 4	Pearson's R = 0.138	43	No correlation
	GPA 3 -> CGPA 4	Pearson's R = 0.164	42	No correlation

4.3 Infographic

The efforts in developing infographics is still at an early stage. The use of analytics software would prove to be a more convenient way to visualize data, and offers real-time data visualization if coupled with machine learning.

5.0 DISCUSSION

The depiction of trends (Intake, Registration Status Profile, and factors (predicted to influence learner retention and academic performance) has been shown here. Nevertheless, the efforts can be improved using machine learning to enable real time data visualization that can offer numerous possibilities. Examples include: (i) evidence-based decision making, (ii) early detection of possible challenges, (iii) track the effectiveness of the interventions introduced to resolve any challenges that is faced by the institution. After initial data cleaning (which was done at the source of data), a second batch of data was extracted to develop an auto-error detection which will be tested using machine learning at a later stage.

Analysis of factors and the study of correlation between the dependent variables (learner retention, and academic performance) and independent variables (demographic and entry qualifications) is proven to be a challenge, but had offered the university various opportunities apart from the improvement of learners' database. Firstly, the richness of the data, and modelling challenges offers great opportunity for learning and capacity building of essential future skills (Programming, Machine Learning, and Data Science). Secondly, the acquired skills would enable the expansion on the research in area of learning analytics (where the factors evaluated are expected to be more dynamic). A concurrent project that concentrates on capacity building in the development of effective (i) pedagogical processes, (ii) learning analytics, and (iii) education learning materials (products) is ongoing. Effort has also been initiated in enhancing the current learning platform using machine learning and artificial intelligence. The third opportunity lies in the beauty of modelling skills which allows the use of the acquired skills to be deployed in various disciplines, thus supports the promotion of interdisciplinary research.

The development of infographics which is still at its infancy stage can benefit from the investment in the analytics software which is targeted for the purpose of modelling. Nevertheless, an important process in this effort is multiple stakeholders' engagement that would ensure effective use of the development infographics. As such, the inclusion of other research methods will still be pursued. The choice of research method must be driven by the objectives of a research.

6.0 CONCLUSION

The exploration in ‘big data’ at the Open University Malaysia had helped to depict trends in Intake, Registration Status Profile, and Factors (predicted to influence learner retention and academic performance of learners). Meanwhile, analysis of dependent variables (learner retention, and academic performance), correlations between the dependent variables and independent variables (demographic and entry qualifications factors) proven to be a challenge. A consequence of the numerous challenges had limited the development of infographics. The paper offers an insight to the transformation of institutional research using the project on learners’ retention and academic performance as a case. The challenges faced in the original study offered the university various opportunities. The use of machine learning can enable real time data visualization that can offer numerous possibilities: (i) improvement in data structure, (ii) evidence-based decision making, (iii) early detection of possible challenges, (iv) tracking of the effectiveness of the interventions, (v) predictive models (vi) capacity building in future skills, (vii) exploration in learning analytics, (viii) evaluation of the efficiency and effectiveness of innovations in pedagogical processes, learning environment, and education learning materials, (ix) promotion of interdisciplinary research, and (x) development of real-time infographics. Such concentrated effort is needed in ensuring quality digital education.

Acknowledgement
This work is funded by the university grant (Grant No: OUM-CR-2019-001). Contributions and the relentless support from Abdul Rauf Ab Rahman (ARU) and Sharifah Syakinah Sy. Dahalan (GICTS) are much appreciated.

References

Bilquise, G., Abdallah, S., and Kobbaey, T. (2020). Predicting Student Retention Among a Homogeneous Population using Data Mining. Proc. of the Int. Conf. on Advanced Intelligent Systems and Informatics. pp. 35-46.

Cerzone, K. (2008). Characteristics of Adult Learners with Implications for Online Learning Design. *AACE Journal*. Vol. 16(2), pp. 137-159.

Crosling, G., Heagney, M., and Thomas, L. (2009). Improving Student Retention in Higher Education: Improving Teaching and Learning. *Australia Universities Review*. Vol. 51, No. 2, pp. 9-18.

Goitsemang, L., and Seretse, T.E. (2019). Evaluating the Impact of using Digital Technologies to Support Open and Distance Postgraduate Learners at Botswana Open University. Proc. of the 9th Pan-Commonwealth Forum, 9-12 September 2019, Edinburgh, Scotland.

Glew, P.J. and Salamonson, Y. (2019). Relationships between Academic Literacy Support, Student Retention and Academic Performance. Nurse Education in Practice. Vol. 39. Pp. 61-66.

Lougheed, P., Drinkwater, A., and Jamieson, L. (2018). Opportunities and Challenges in Predictive Modelling for Student Retention. A Report by The Higher Education Quality Council of Ontario, Canada.

Perry, S.R., and Sullivan, R.B. (2007). Student Affairs, Retention and the Enrollment Management Dynamics. American Association of Collegiate Registrars and Admissions Officers. (http://www4.aacrao.org/semsource/sem/indexc802.html?fa=view&id=3506)

Powell, S., and MacNeill, S. (2012). Institutional Readiness for Analytics. CETIS Analytical Series. Vol. 1, No. 8.

Shavkun, I, Bukharina, L., Dybchynska, Y., and Onyshcchenko, O., (2021). Social Economic Factors of ICT Use in Education: Lessons from the Pandemic. ICTERI 2021. Pp. 193-203.

Tresman, S. (2002). Towards a Strategy for Improved Student Retention in Programme of Open, Distance Education: A Case Study from the Open University UK. *The International Review of Research in Open and Distributed Learning*. Vol. 3, No. 1.

Tung, Lai Cheng. (2012). Proactive Intervention Strategies for Improving Online Student Retention in a Malaysian Distance Education Institution. *MERLOT Journal of Online Learning and Teaching*. Vol. 8, No. 4.

World Economic Forum. (2020). Accelerating Digital Inclusion in the New Normal. July 2020 Playbook by World Economic Forum and Boston Consulting Group.