Developing Predictors for Student Involvement in Generic Competency Development Activities in Smart Learning Environment

Joseph C. H. So
College of Professional and Continuing Education,
The Hong Kong Polytechnic University
joseph.so@cpce-polyu.edu.hk

Ho Wai Tung
College of Professional and Continuing Education,
The Hong Kong Polytechnic University
ccwtho@cpce-polyu.edu.hk

Ada Chan
College of Professional and Continuing Education,
The Hong Kong Polytechnic University
ccada@cpce-polyu.edu.hk

Simon Wong
College of Professional and Continuing Education,
The Hong Kong Polytechnic University
ccswong@cpce-polyu.edu.hk

Adam Wong
College of Professional and Continuing Education,
The Hong Kong Polytechnic University
spklwong@cpce-polyu.edu.hk

Henry Chan
Department of Computing
The Hong Kong Polytechnic University
henry.chan.comp@polyu.edu.hk

***Abstract*— Smart Learning Environment (SLE) aims at promoting personalized education in various form with different settings fitting the learners' needs. Many works have done to realise the environment for academic studies. However, the development of the generic competencies is another key element of education. The engagement of students in the developmental activities of generic competencies (GDA) is a main concern of the organizers in tertiary education institutions, particularly the student affair offices. They want to have some predictors to reflect the participation of students with some identifiable factors so that the provision can planned correspondingly. In this work, we attempted to the evaluation on a set of attributes of students to their participation on the GDA by means of the correlation and the classification through logical regression. We studied the records of 1649 graduates in a tertiary education institution across two academic years and found that some single factors are reliable in predicting the tendency of students in taking part in the GDA.**

Keywords—Generic competency; Smart Learning Environment; Logical Regression; sub-degree education; co-curricular activities; extra-curricular activities; student affairs

I. Introduction

The Smart Learning Environment (SLE) has become known as a promising area of incorporating various technologies and enhancing interactivity and personalisation in the learning environment. It is also advocated that SLE not only allows users to interact with the learning system at their convenience, but can also 'actively provide that necessary learning guidance, hints, supportive tools and learning suggestions' [1]. Much discussion has revolved around the development of SLE in teaching and learning [2] [3] [4] [5]. The smart pedagogy framework [2] is a four-tier model on recent technological advances promote the smart pedagogies for learners to be equipped from knowledge and core skills to collective intelligence. Kinshuk [3] further highlighted the growing difference of the prior knowledge among students, and it implies the need of more personalised teaching. Martens's team. [4] offered insight into how smart systems support teaching and learning, based on the experience of MeinKosmos in a German university and Zhuang's team connected the learning environment to the formation of smart cities and the relationships of different learning environments [5].

The personalised learning experience is a key feature imported into SLE. To enhance personalisation of the learning platform, previous studies have proposed the use of various technologies to enhance LMS and support student learning in the formal curriculum and discipline-oriented subjects. With the evolution of enabling technology, intelligence can be more thoroughly incorporated in LMS to develop systems that personalise adaptation. In adaptive education systems, such advances address the individual level by analysing students' individual personality traits and skills and by adapting to individual needs during the delivery of learning content and instruction [6].

Many studies have examined the use of technology to reinforce learning, such as the Mobile Learning Support System (MLSS) for ubiquitous learning [7], and on systems compatible with LMS that incorporate self-regulated learning and intelligence [8] [9] [10] [11] . In Kumar's team development [8], a framework called Smart Competence Analytics and LEarning (SCALE) attempts to use learning traces to extract the learners' underlying competence level and incorporate context-aware and personalised learning into mass-marketed LMSFurthermore, the incorporation of intelligence into LMS is a promising area. For example, the intelligent tutoring system (ITS) [9] is built on a multi-agent system to provide students with tutoring that is adaptive. Bajaj and Sharma [10] reviewed several AI techniques cab potentially added in adaptive education systems, including Decision Tree, Hidden Markov Models and Generic Algorithms, which use multilayer perceptrons and decision trees to determine students' learning styles. Smart LMS are expected to link smart devices with embedded AI systems [12].

The education of generic competencies (GC) is critical in the whole person development and the significant of generic competence is growing in recent years. Therefore, in the SLE,

978-1-7281-7303-0/20/$31.00 ©2020 IEEE
DOI 10.1109/COMPSAC48688.2020.00287

the education should be limited to academic studies. In the future SLE, it is expected that the education will not only support the learning in formal curriculum or the skill development. The environment should also nurture the whole person development if it becomes a fully-fledged ecosystem. In a work in developing an i-campus, it also concurs that inclusion of extra-curricular activities is an advantage to personalized learning [13].

Generic competency development activities (GDA), including co-curricular and extra-curricular activities, aims to develop students GC. There are some works trying the correlate the involvement in extra-curricular activities to the academic results. The effect of involving extracurricular activities undertaken in high school in math achievement [14] as well as whether the better performance in clinical training due to their extracurricular activities is studied [15]. Many of these efforts focus of the impact of the GC development activities to the academic performance. However, as GDA are part of the education, the student engagement is also key concern in a comprehensive learning environment.

The prediction of the student behavior is very important for the educators, in particular to the student affairs professionals, in designing meaningful and well received development activities for the students [16]. Such foreseeing feature should be found in the future SLE. Many works tried to predict the student performance with different metrics[17], [18]. In [18], 19 indicators are introduced to the predict the success of student in an online learning platform in China. However, there are very little work on the prediction on the student engagement. Can we have some sort of the indicators to predict the involvement of student in GDA? In our previous work, we have studied on the difference of STEM and Non-STEM in participation rate in GDA, we found that students from STEM programs tend to be less involved in the activities [19]. However, it would be very useful for the organizers of the activities to have better understand on the pattern and factors affecting the students' intention and motivation in the taking part in the activities.

In the current work, we attempt to explore the way to predict the student engagement in generic competency development activities (GDA) from the factors of the students' information including the public examination results in secondary school, the academic program and its structure, as well as the academic results at the graduation from the college.

II. Methodology

The data are collected from a post-secondary institution in Hong Kong which offers sub-degree programs including, Associate Degree (AD) and Higher Diploma (HD) programmes, spanning the domains of arts, science, social sciences, business and the areas of design and health studies for secondary school leavers. The institution has the largest portion of sub-degree students in Hong Kong.

The samples are those who have taken surveys on their generic competencies in the final semester of the studies and the surveys are conducted in 2016 and 2017. The data are gathered on their students academic results in the College, the survey on their public examination results before the admission to the College, the records of the participation on in the reports for each of students, named Cocurricular Achievement Transcript (CAT) and their demography information including the gender, the program of studies and the program structure. CAT records the actual involvement of each student in every GDA, including co-curricular and extra-curricular activities, one has participated in the actual duration of the activities.

III. Results and Findings

A. About the collected data

Table II summarizes the sample size students in each of the year of graduation. There are 929 and 727 students taken as samples in 2015 and 2016 cohort respectively.

The students from different programs of studies are grouped into STEM or non-STEM programs in the college. The STEM group includes associate degree students in Engineering, Information Technology, Science, Health Studies, Statistics and Computing for Business and Higher Diploma students in Mechanical Engineering.

The non-STEM students in all other programs are includes those who are studying programs in the areas of social sciences, humanities, design and business.

TABLE II. Characteristics of Sample Data

	Number of Students	
	2015 Graduates	***2016 Graduates***
Academic Discipline		
STEM	266 (28.63%)	186 (25.58%)
Non-STEM	663 (71.37%)	541 (74.42%)
Program Structure		
AD	874 (94.08%)	689 (94.77%)
HD	55 (5.92%)	38 (5.22%)
Gender		
F	569 (61,25%)	424 (60.80%)
M	360 (38.75%)	285 (39.20%)
Total	929 (100%)	727 (100%)

In general, there is an average of 15.94 hours of participation of a students with standard deviation of the 16.8, and the 5 hours for the 25 percentile and 20.5 hours for 75% percentile. In 2015, the average participation during is 15.72 hours and 16.22 hours in 2016. The distribution of the two cohorts are quite similar.

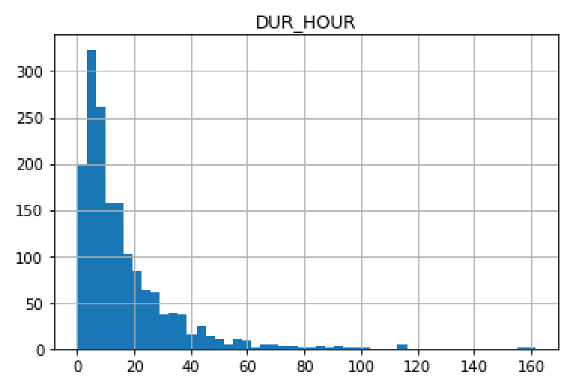

Figure 1 Distribution of the participation hours

B. Findings

1) By Gender

As shown in TABLE III, it shows that average participation hours of female being 17.3 hours is higher than that of male being 13.5 hours in 2015 and 17.12 hours for female and 14.99 hours for male in 2016.

Also, it is further noted that for the STEM students the difference is even greater, 15.89(F) to 11.00(M) in 2015, even though the gap is not that large in 2016, 13.26(F) to 12.32(M).

It noted that female students are more participative than the male regardless whether they are studying STEM programmes.

TABLE III. PARTICIPATION OF GENDER AND STEM

Gender	Duration of Participation in Number of Hours – Mean (Std. Dev.)	
	2015 Graduates	2016 Graduates
STEM		
F	15.89(15.3)	13.26 (14.67)
M	11.00(14.13)	12.32 (13.67)
non-STEM		
F	17.65(16.97)	18.04 (18.85)
M	15.17(14.75)	16.46 (19.32)
Overall		
F	17.3 (16.62)	17.12 (18.21)
M	13.5 (14.63)	14.99(17.6)
Total	15.72 (15.91)	16.22(17.95)

2) By Academic Background

We attempt to investigate the correlation of the academic performance to the involvement of the whole personal development activities. We study two key aspects of the academic performance: before college studies, and after College studies.

Over 95% students in the College have taken Hong Kong Diploma of Education Examination (HKDSE) in the final year of the studies in secondary school. The results are good indicators the academic performance before enrolling to the College. As students have taken different elective subjects, for the fair comparison of their results, we took the four core subjects to be taken by over 98% students, including Chinese (CHI), English (ENG), Mathematics (MATH) and Liberal Studies (LIB). Then, we correlate the grades in these subjects to the duration of participation in activities in the whole period of studies in the College.

TABLE IV. Correlation of Participation to HKDSE Results.

Subject	Correlation	
	2015 Graduates	*2016 Graduates*
CHI	-0.055484	*-0.015612*
ENG	0.026027	0.034629
MATH	0.008757	0.018188
LIB	0.024443	-0.001347
DSE4Core	-0.000316	0.015285

From the Table IV, it shows that the correlation of "DSE4Core" to "DUR_HOUR" is less than 0.01. Even we looked into a particular subject, for each of CHI, ENG, MATH and LIB, the correlation values are very low even though both positive and negative values are found.

3) Academic Program Structure (AD vs HD)

In section, we aim at investigating the effect of the academic structure of the studies. The structure of associate degree (AD) programs have 60% of general education (GE) subjects and 40% of discipline specific (DS) subjects while that of higher diploma (HD) have 40% of GE subjects and 60% of DS ones. We investigate if the students studying more GE subjects tends to more willing participating in GDA.

As shown in Figure 2 Participation Hours in Academic Structure, the average of duration of participation of students from AD programs are more than that of HD. However, it is noted the most of the students, 1563 or 94.4%, are AD students. Further work is needed to investigate the difference of the two groups.

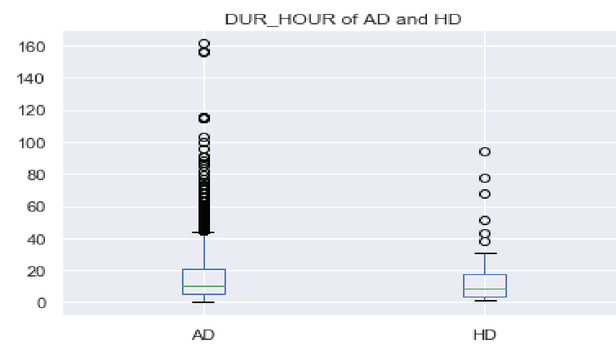

Figure 2 Participation Hours in Academic Structure

4) Academic Program

The academic programs in the College are grouped into eight schemes (ProgCode), namely, Applied Social Sciences (8C111), Business (AD) (8C108), Business (HD) (8C121), Design (8C113), Health Studies (8C106), Humanities and Communication (8C110), Mechanical Engineering (HD) (8C118), and Science and Technology (8C112), The distribution of the participation is plotted in Figure 3. The results show that students from the programmes of Applied Social Sciences and Humanities and Communication have more hours of involvement in GDA.

5) Academic performance in the College

We also consider the students' academic results at the completion of the studies. After the college studies, we look into the academic results in terms of grade point average (GPA) in the Semester GPA of the final semester and the overall Communitive Grade Point Average(CumGPA) as the graduation.

The correlation values are around 0.08 for the both years in Semester GPA and slight greater than 0.1 for CumCPA for the both cohorts.

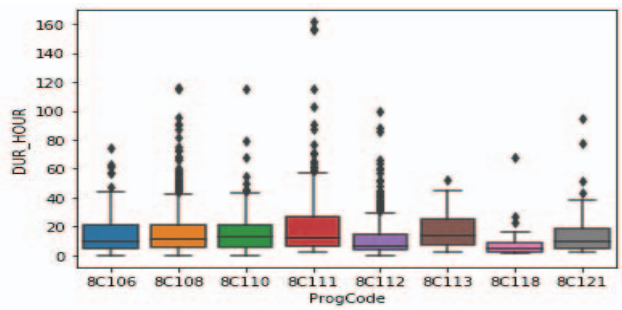

Figure 3 Participation Hours in Academic Programs.

It shows that the correlation is very low. On one hand, it shows the academic results is not a predictor on the involvement in GDA. One the other hand, it also shows that the engagement in GDA do not affect neither positively nor negatively on the academic performance.

TABLE IV. Correlation of Participation to HKDSE Results.

GPA	Correlation		
	2015 Graduates	2016 Graduates	Two Cohorts
Final Semester GPA	0.082328	0.081766	0.081586
Overall	0.118877	0.132883	0.125795

In sum, it is obvious that there very little correlation among the involvement and the academic results, on both before and after the studies in the College.

It is obvious that there very little correlation among the involvement and the academic discipline of studies. From the above findings, it shows that the correlation of "DUR_HOUR" to "Gender", Schemes, Program, and HD/AD are very low.

C. Classification by Logistic Regression

In addition to simple correlation, we further attempted to the perform classification through logistic regression. The attributes used includes the academic program code (ProgCode), four core subjects in HKDSE (including Chinese (CHI), English (ENG), Mathematics (MATH) and Liberal Studies (LIB)) and their summation, the academic results in the College (CumGPA), Gender, whether in science discipline (isScience) and program structure (HD or AD). The dependent various is the duration of the participation in hours (DUR_HOUR).

We run the Logistic Regression in Scikit-learn in python environment. We classified the DUR_HOUR greater than 10 hours to high and less than that to be low. To ensure the internal validity, we split the data into two set: training set and prediction set. 80% of randomly selected data are used as input for training the algorithm and build the mode. The resting 20% are used is evaluation.

1) Confusion matrix

As shown in Figure 4 Confusion Matrix of the classifier performance, the first row is for students whose actual participation is high, that is in the group of 1. Out of 330 students in prediction set students, the 145 of them are actually having high participation. And out of then, the classifier correctly predicted 118, while for remaining 37 students, the classifier predicted those 0, which is not correct. There are 25.5% (37/145) predicted incorrectly.

At the second row, there were 175 students whom their participation is low, that is the group of 0. The classifier correctly predicted 74 (42.3%) of them, and the remaining wrongly as 1. So, it did not give a good prediction for the students with low participation.

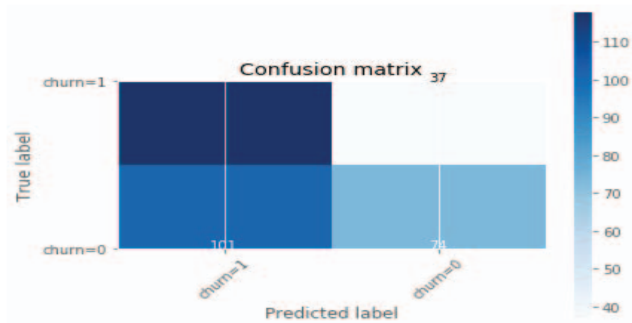

Figure 4 Confusion Matrix of the classifier performance

TABLE V Classification Performance

	Precision	Recall	F1-score	Support
0	0.67	0.42	0.52	175
1	0.54	0.76	0.63	155
accuracy			0.58	330
macro avg	0.60	0.59	0.57	330
weighted avg	0.61	0.58	0.57	330

We further calculate the precision and recall of each label, where precision is a measure of the accuracy provided that a class label has been predicted. It is defined as:

$$Precision = TP / (TP + FP)$$

and recall is true positive rate. It is defined as:

$$Recall = TP / (TP + FN)$$

where TP is true positive, FP is false positive and FN is false negative.

The F1 score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0. It is a good way to show that a classifier has a good value for both recall and precision. F1 score is 0.58 which is an average accuracy.

IV. IMPLICATIONS OF RESULTS AND CONCLUSIONS

We have evaluated some factors that are commonly used as the significant information about students before they enter to the institution to the time when they complete the studies. It is observed there is no single factor can be used with confidence to predict the participation in GDA, even with logical regression on the combinations of these factors.

Knowing that it is very useful for the student affairs office to understand the pattern of students' involvement in GDA, the development of SLE has to gather other data sources and the other ways of the data processing to the achieve some more in depth understanding on the pattern so that some sensible prediction can be obtained. There are some other directions worth of the further consideration. Students have different personalities, so they have different interests. Can information of the personality be incorporated into the prediction in SLE? Furthermore, can the prediction also include the feedback on the activities after they have participated or even if they have chosen not to joined? These data are not well structured. Some semantic analysis on the data gathered in SLE are needed to proceed with these directions.

The impact of the heavily engagement in non-academic activities is a hot topic in the area of student development. There are some controversy whether there are positive or negative effect on academic results. the more involvement the better, the less involvement the better, or good to involve in some extent, but bad after some threshold [20].

Furthermore, from our results, on the effect of the academic results, the low positive correlation has shown the academic performance cannot be used to predict the involvement in GDA, but it is positive, therefore, it is safe to say the engagement in GDA does not lead to negative impact the academic results. It is important for the practitioners in student affairs and the tertiary institutions that encouraging students to participates in GDA will be leading to decline in academic results. Some further studies is needed if some threshold model be considered.

V. References

[1] G. J. Hwang, "Definition, framework and research issues of smart learning environments - a context-aware ubiquitous learning perspective," *Smart Learning Environments,* vol. 1, no. 4, 2014.

[2] Z. T. Zhu, M.-H. Yu and P. Riezebos, "A research framework of smart education," *Smart Learning Environments,* vol. 3, no. 1, pp. 2196-7091, 2016.

[3] Kinshuk, N.-S. Chen, I.-L. Cheng and S. W. Chew, "Evolution Is not enough: Revolutionizing Current Learning Environments to Smart Learning Environments," *International Journal of Artificial Intelligence in Education,* vol. 26, no. 2, pp. 561-581, 2016.

[4] A. Martens, K. Sandkuhl, B. Lantow, H. Lehmann, W. D. Lettau and F. Radisch, "An evaluation approach for smart support of teaching and learning processes," *Smart Learning Environments,* vol. 6, p. 2, 2019.

[5] R. Zhuang, H. Fang, Y. Zhang, A. Lu and R. Huang, "Smart learning environments for a smart city: from the perspective of lifelong and lifewide learning," *Smart Learning Environments,* p. 4, 2017.

[6] K. Almohammadi, H. Hagras, D. Alghazzaw and G. Aldabbagh, "A Survey of Artificial Intelligence Techniques Employed for Adaptive Educational Systems within E-Learning Platforms," *Journal of Artificial Intelligence and Soft Computing Research,* vol. 7, no. 1, pp. 47-64, 2017.

[7] K. Y. Chin and Y. Lin Chen, "A Mobile Learning Support System for Ubiquitous Learning Environments,," *Procedia - Social and Behavioral Sciences,* vol. 73, pp. 14-21, 2013.

[8] V. Kumar, D. Boulanger, J. Seanosky and a. et, "Competence analytics," *Journal of Computers in Education,* vol. 1, no. 4, pp. 251-270, 2014.

[9] C. Giuffra P, R. Silveira and M. Nakayama, "An Intelligent LMS Model Based on Intelligent Tutoring Systems," in *12th Int'l Conference on Intelligent Tutoring Systems*, Honolulu, 2014.

[10] R. Bajaj and V. Sharma, "Smart Education with artificial intelligence based determination of learning styles,," in *Procedia Computer Science*, 2018.

[11] K. Colchester, H. Hagras, D. Alghazzawi and G. Aldabbagh, "A Survey of Artificial Intelligence Techniques Employed for Adaptive Educational Systems within E-Learning Platforms," *Journal of Artificial Intelligence and Soft Computing Research,* vol. 7, no. 1, pp. 47-64, 2017.

[12] mylearningspace, "Smart LMS: AI and The Future of eLearning," 4 Jun 2018. [Online]. Available: https://www.mylearningspace.com.au/news/smart-lms/. [Accessed 20 Jan 2019].

[13] L. f. Kwok, "A vision for the development of i-campus," *Smart Learning Environments,* vol. 2, no. 2, 2015.

[14] D. S. Morris, "Extracurricular Activity Participation in High School: Mechanisms Linking Participation to Math Achievement and 4-Year College Attendance," *American Educational Research Journal,* vol. 53, no. 5, 2016.

[15] L. Urlings-Strop, A. Themmen and K. Stegers-Jager, "The relationship between extracurricular activities assessed during selection and during medical school and performance.," *Adv Health Sci Educ Theory Prac,* vol. 22, no. 2, pp. 287-298, 2017.

[16] Kuh, "What we're learning about student engagement from NSSE," *Change,* vol. 35, no. 2, pp. 24-32, 2003.

[17] A. U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," in *IOP Conference Series: Materials Science and Engineering*, 2017.

[18] W. Zhang, X. Huang, S. Wang, J. Shu, H. Liu and H. Chen, "Student Performance Prediction via Online Learning Behavior Analytics," in *2017 International Symposium on Educational Technology (ISET)*, Hong Kong.

[19] W. T. Ho and J. C. H. So, "Issues in Measuring Generic Skills Using Self-Administered Questionnaire in a Community College in Hong Kong," *Open Access Library Journal,* pp. 1-8, 2016.

[20] P. Buckley and P. Lee, "Contents Full Article Content List Abstract The modern higher education context Academic performance Competence and skill development Methodology Results Discussion and conclusion References Article Metrics Related Articles Cite Share Request Permiss," *Sage Journals,* 31 10 2018.