International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Profiling Students in the K to 12 Curriculum Strands Based on Academic Performance Using Cluster Analysis

Pierre Jared T. Morados
Science and Mathematics Education Department, University of San Carlos, Cebu City, Philippines-6000
Email address: morados.etsmed @ gmail.com

***Abstract*— *The addition of the senior high school in the K to 12 Curriculum is intended to help students prepare for their future career. By choosing a specific strand, the students are equipped with knowledge and skills that they would need in their desired field; hence, students in different strands are expected to be taking up different courses throughout their senior high school. This also means that in order to have academic success, students' abilities must fit their chosen strand. The current study aims to categorize students based on their academic performance to determine the type of students in each strand of the K to 12 Curriculum. Specifically, it employs k-Means clustering to categorize students based on their Earth Science and General Mathematics scores. As the results showed, there were four clusters identified, which reflected the abilities of students in each cluster. Moreover, MANOVA revealed that these four clusters are significantly different from each other in both Earth Science and General Mathematics scores. Further, Chi-square test uncovered the significant association between the clusters and the students' strand. It was then found out that STEM students' abilities were fit to their chosen strand. However, the same cannot be said for the other strands since only the students' Earth Science and General Mathematics scores were used in the study. Hence, further research must be done that would include other measures of students' performance.*

***Keywords*—*Academic Strands:Cluster Analysis:K to 12 Curriculum:Student Profiling.*

I. Introduction

One of the several changes that the K to 12 Curriculum brought in the Philippine educational system is the addition of the senior high school in its basic education. Referred to as the apex of secondary education, the senior high school prepares students for their desired career paths [1]. Here, students choose from a set of educational strands – each designed to equip them with knowledge and skills required for their future professions – to specialize.

This means students in a particular strand have to take up courses different from those in the other strands. For instance, students from the Science, Technology, Engineering, and Mathematics (STEM) strand go through relatively more science and mathematics subjects to prepare them for future science and engineering professions; while students from Accountancy, Business, and Management (ABM) strand have to take up business courses, preparing them to be future entrepreneurs; students from Humanities and Social Sciences (HUMS) strand are offered philosophy and political science courses suited to their field; and students from technical-vocational strands (e.g. Bread and Pastry, and Animation) are exposed to courses training them to be competent in their chosen vocations [2]. Hence, in order to be successful in their studies, students' abilities must also fit to their chosen strand.

Cluster analysis is a set of multivariate techniques used to group cases according to characteristics they possess [3]. This analysis has become popular in the field of education since it has been a tool usually utilized in student profiling and mapping. Several studies [4][5][6] have applied cluster analysis to categorize students based on their academic performance. In particular, these studies commonly employ k-Means Cluster Analysis – a non-hierarchical cluster analysis known for its ease and simple implementation, and ability to accommodate sparse data [4].

II. Research Objectives

This study intends to categorize students based on their academic performance, as measured by their Earth Science and General Mathematics scores, to determine the type of students in each track of the senior high school. Specifically, it aims to answer the following research questions:
1. Is the data appropriate for multivariate analyses?
2. How can the students be clustered according to their Earth Science and General Mathematics scores?
3. Is there a significant difference in the students' Earth Science and General Mathematics scores among the clusters?
4. Is there a significant association between the students' clusters and strands?

III. Methodology

Data of 285 grade 11 senior high school students from a public high school in Cebu, Philippines were utilized in the study. Particularly, they represent each of the strand – STEM (N = 75), ABM (N = 59), HUMS (N = 50), Bread and Pastry (N = 58), and Animation (N = 43) – offered in the said school.

Before the data was gathered, permission from the school's principal, as well as the advisers, was sought. Records of the students were then collected from their respective advisers, who decided to cooperate in the conduct of the study. Particularly, the records are the teachers' School Form 14, which reflected the students' first quarterly examination scores. From this form, the students' Earth Science (ES) and General Mathematics (GM) scores were extracted and were used in the study. These two courses are core subjects that all of the students, regardless of their strands, have to take. The quarterly examination is a single paper-pencil test, collectively made by all subject teachers; which included similar items across the different strands, covering the competencies listed in the curriculum guide for the first quarter. Moreover, it constitutes 25% of the students' quarterly grade in these two core subjects. For this reason, the quarterly examination scores in ES and GM were utilized in the study, as measures of the students' performance.

Prior to the application of the multivariate analyses (MVAs), exploration of the raw data was done. This is to examine the presence of missing values and outliers, and make decisions to remedy them. In addition, other tests of assumptions (e.g. tests of normality and collinearity) were conducted.

After the tests of assumptions were done, cluster analysis of the students based on their ES and GM scores commenced. K-Means Cluster Analysis was used in clustering, with four predicted clusters. Results of this cluster analysis were then subjected to Multivariate Analysis of Variance (MANOVA) to verify the validity and strength of the clustering. This was done by determining whether there is a significant difference in the students' ES and GM scores when grouped according to the identified clusters. Finally, Chi-square test was done to determine if there is a significant association between the identified clusters and strands.

IV. RESULTS

A. Data Exploration

Inspection of the raw data revealed that out of the 285 cases, there were three data missing for both ES and GM scores. These missing values belonged to the same cases.

Box plot analyses were utilized to detect outliers. Results showed that there were no outliers in the students' GM scores; however, there were three in ES. Further analysis revealed that these same cases were also outliers when students are grouped with respect to their strands.

Collectively, cases with missing data and outliers only constitute 2.11% of the total. According to Garson (2015) these may be dropped; hence, the said cases were not included in further analyses.

B. Descriptive Statistics

After all the cases with missing data and outliers were discarded, 279 cases were left for cluster analysis. Table I shows the mean and standard deviations of the cases' ES and GM scores.

TABLE I. Mean and Standard Deviation of Students' Earth Science and Mathematics Scores

	N	*M*	*SD*
Earth Science	279	31.46	10.52
General Mathematics	279	35.09	7.19

C. Test for Assumptions

Using Kolmogorov-Smirnov normality test, it was found out that both the students' ES, $D(279) = 0.10$, $p < 0.01$, and GM, $D(279) = 0.07$, $p < 0.01$, scores are normally distributed. This satisfies the normality assumption of the MVAs to be performed.

Moreover, Pearson correlation was performed to determine if there is a significant relationship between the students' ES and GM scores. As the results suggest, there exists a significant relationship, $p < 0.01$, between these two variables; that is, they are positively moderately correlated, r = 0. 59. Despite this correlation, the data satisfy the multicollinearity assumption ($r < 0.6 - 0.8$) of the dependent variables.

D. Clustering Students Based on ES and GM Scores

The students were grouped into four clusters using k-Means Cluster Analysis. Table II shows the descriptive statistics of these clusters. Note that the data are not standardized since they only represent one measure, which is scores. Moreover, the highest possible scores of the quarterly examinations in all subjects were all set to 50; hence, it is not necessary to express these values in z-scores.

TABLE II. Mean and Standard Deviations of Students' ES and GM Scores within Each Cluster

Cluster	*N*	**Earth Science**		**General Mathematics**	
		M	*SD*	*M*	*SD*
1	79	36.19	3.53	24.67	4.31
2	52	25.17	5.13	17.60	4.94
3	70	32.94	4.23	36.91	4.31
4	78	42.51	3.61	42.67	4.05

Based on the descriptive statistics presented in table II, the characteristics of students in each cluster may be identified. Table III summarizes these clusters and characteristics.

TABLE III. Characteristics of Students in Each Cluster

Cluster	**Characteristics**
1	Relatively better-performing in ES than in GM
2	Low-performing in both ES and GM
3	Relatively better-performing in GM than in ES
4	High-performing in both ES and GM

E. Comparison of Students' ES and GM Scores Among Clusters

The Box's test of Covariance results revealed that the homogeneity-of-covariance-across-the-group assumption was not met, Box's M (18.98), $p = 0.03$. Using Pillai's Trace criterion, it was found out that there is significant difference, $F(2, 274) = 1.27$, $p < 0.01$, among the clusters in both ES, $F(3,275) = 446.27$, $p < 0.01$, and GM, $F(3,275) = 197.91$, $p < 0.01$, scores.

Moreover, post hoc analysis using Duncan's Test uncovered that each of the clusters are significantly different from any other cluster. Thus, this confirms the validity and strength of the identified clusters.

F. Relationship Between the Students' Clusters and Strands

The Chi-square test revealed that there is a significant association, X^2 (12, N = 279) = 300.23, $p < 0.01$, between the identified clusters and the strands. As table IV shows, majority of the STEM students belong to cluster 4; ABM students are in clusters 1 and 4; HUMS students in cluster 1, Bread and Pastry in cluster 3; and Animation students in cluster 2.

TABLE IV. Percentage of Students Belonging to each Cluster within Strands

CLUSTER	STRAND				
	STEM	ABM	HUMS	BREAD AND PASTRY	ANIMATION
1	4.1	47.4	66.0	21.4	9.5
2	0	12.3	4.0	12.5	85.7
3	25.7	12.3	12.0	64.3	4.8
4	70.3	28.1	10.0	1.8	0
TOTAL	100	100	100	100	100

V. DISCUSSION

The aim of this study is to categorize students according to their academic performance to identify the type of students in each strand of the senior high school. As the results suggest, there were four significantly different clusters identified when students are categorized based on their Earth Science and General Mathematics scores. Each of these clusters represents the abilities of the students who belonged in them; that is, Cluster 1are students who are relatively better-performing in ES than in GM, Cluster 2 are students who are low-performing in both ES and GM, Cluster 3 are students who are relatively better-performing in GM than in ES, and Cluster 4 are students who are high-performing in both ES and GM.

It was also found out that there was a significant association between the identified clusters and the students' strands. Results revealed that majority of the students in the i) STEM strand are in cluster 4, ii) ABM strand are in clusters 1 and 4 iii) HUMS are in cluster 1, iv) Bread and Pastry strand are cluster 3, and v) Animation strand are in to cluster 2.

VI. CONCLUSION

The findings of the study reflect the type of students in every strand. That is, STEM students excel in both ES and GM; ABM students are average-performers in both ES and GM, HUMS students are those who are better in ES than in GM; Bread and Pastry students are those who bettered in GM than in ES; while Animation students are low-performing in both ES and GM. Based on these descriptions, it can be said that there is no mismatch in the STEM students' abilities and their chosen strand. These students, who need to be good in both Science and Mathematics subjects, are actually so. However, since the study only utilized the students' ES and GM scores, it is not possible to say the same with the other strands since subjects directly related to these strands were not involved in the study.

With this regard, further studies must be done. The current study only focused on the ES and GM quarterly examinations scores to cluster the students. Obviously, there are other indicators of students' performance that could serve as measures of their abilities, and be used as variables in clustering the students. This would make the results more comprehensive and better capture reality. However, this study can serve as a guide for future related researches, such as adapting its methodology or replicating it in other contexts.

REFERENCES

[1] Department of Education, "The K to 12 Basic Education Program (as of March 12, 2012)", [online] Available at http://deped.gov.ph, 2012.

[2] Department of Education, "Academic Track", [online] Available at https://www.deped.gov.ph/k-to-12/about/k-to-12-basic-education-curriculum/academic-track/,ND.

[3] J. F. Hair, W. C. Black , B. J.Babin, R. E. Anderson, &R. L. Tatham, "Cluster Analysis" in *Multivariate data analysis: Pearson New International Edition*, 7th ed., pp. 415-474, 2014.

[4] O.J. Oyelade, O.O. Oladipupo, &I.C. Obagbuwa, "Application of k Means Clustering algorithm for prediction of Students Academic Performance", 2010

[5] O.N. Darcan, &B.Y. Badur, "Student profiling on academic performance using cluster analysis", *Journal of e-Learning & Higher Education, 2012*.

[6] A. P. Alfiani, &F. A. Wulandari, "Mapping student's performance based on data mining approach (a case study)", *Agriculture and Agricultural Science Procedia*, issue *3*, pp. 173-177, 2015.

[7] G. D. Garson, (2015), "Missing values analysis & data amputation (2015 ed.)", *Asheboro, NC: Statistical Associates Publishers*, 2015.