Webology (ISSN: 1735-188X)
Volume 15, Number 1, 2018

Developing A Prediction Model For Academic Success Of Physics Major Students: An Analysis Employing Data Mining Technique

Vinchail A. Siason

Western Mindanao State University, Philippines.

Abstract

In this paper, factors that influenced the success of Physics students in a Philippine university was determined. Specifically, the study aimed to find out whether significant correlation exists between the 1st Year grades of the students in Mathematics, Science, English and College Entrance Test (CET) rating with their retention in the BS Physics program. It also generated a predictive model of success for the students and determined which variable is the greatest predictor for the student study outcome. The study employed the descriptive correlation method research design. Data was analyzed using Point-biserial Correlation and Logistic Regression analysis. Results using Point-Biserial Correlation at $p<0.05$ disclosed that all of the cognitive variables (First year subjects and CET Results) were found to be correlated with the student study outcome at varied measures. Binary Logistic Regression analyses generated the prediction model and deduced that the grades in Calculus with Analytic Geometry and English have the most significance in predicting the success of the Students in the BS Physics program in the specific university while the CET Results and Algebra were found to have no influence in predicting the study outcomes of the students. The generated statistical model is found to be valid and acceptable with overall percentage accuracy of 84.8%.

Keywords: Student Retention, Data Mining, Logistic Regression, Physics students

Introduction

Background of the Study

Student retention has long become a concern and issue among academic institutions. Relative to this, Aljohani (2016) claimed that it has become a topic of insterest since the 1600s. To this date, it has become a global issue especially among Higher Education Institutions (HEIs).It is supposed that the reason why student retention has attracted the attention of different stakeholders (e.g., teachers, parents, policy makers, school managers, among others) is the fact that it is a key factor of the survival of academic institutions (Howard & Sharpe, 2019).

Low student retention is an important issue for all university policy makers due to the potential negative impact on the image of the university and the career path of the dropouts. No wonder, increasing student retention or persistence is a long-term goal in all academic institutions. The consequences of student attrition are of considerable concern for students, academic and administrative staff.

The most vulnerable students to low student retention at all institutions of higher education are the first-year students, who are at greatest risk of dropping out in the first term or semester of study or not completing their program or degree on time. Therefore most retention studies address the retention of first-year students (Horstmansh of & Zimitat, 2007; Noble et al., 2007; Strayhorn, 2009). Consequently, the early identification of vulnerable students who are prone to drop their courses is crucial for the success of any retention strategy. This would allow educational institutions to undertake timely and pro-active measures. Once identified, these 'at-risk' students can be then targeted with academic and administrative support to increase their chance of staying on the course. A number of theoretical models have been developed to explain what keeps students on a course. Based on an extensive literature review of dropout in e-learning environment, Jun (2005) identified variables that may impact attrition and have been included in theoretical models of dropout. He classified them into five constructs (i.e., factors: individual background, motivation, academic integration, social integration and technological support).

Knowledge discovery in databases, often called data mining, aims at the discovery of useful information from large collections of data. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making and it has become an essential component in various organizations. There are increasing research interests in using data mining in education to develop methods that discover knowledge from data originating from academic environments. Researchers have also been able to extend student modelling even beyond educational software, towards figuring out what factors are predictive of student failure or non-retention in college courses or in college altogether (Dekker et al. 2009; Romero et al. 2008).

In the Physics department of a particular university in the Philippines, student retention has been one of the most pressing concerns due to the general decline of student enrolees in the succeeding semesters after they have enrolled in their first year. The attrition of Physics students in the program seems to be general across all year levels. Graduation rates of student entrants have been found to be roughly 40-50% according to existing data. The utilization of these data from the previous entrants and graduates of the program is deemed critical to probe into the problem and find solutions to better the interventions in addressing the needs of the students and facilitate better learning for them that would eventually positively influence their retention. According to Tinto (2006), learning has always been the key to student persistence. Students who learn are students who stay. Institutions that are successful in building settings that educate all their students, all not just some, are successful in graduating their students.

Student Success, Attrition, and Drop Outs

For decades now, considerable amount of hard work has been given to understand the process of student retention and graduation in higher education institutions. In congruence to this, numerous researches have been conducted on the determination of factors affecting student success in different academic disciplines. These studies aimed to produce findings that can be utilized by academic institutions and government educational agencies to come up with policies and programs that can lower the attrition rate of students.

A wide range of researches and investigations have already been conducted to establish that student's background/demographics and incoming academic ability (i.e., pre college measures) are important predictors of a student's ability to persist to graduation (Ishitani, 2003; Ishitani & Snider, 2006; Perkhounkova et al., 2006; Tinto, 1975). Previous retention research on students at the University of Minnesota-Twin Cities (UM-TC) found that not only were background and pre-college characteristics important, but also that ''academic fit,'' as measured by admission to a student's first choice college and first-term academic progress, were significant predictors of academic success (Radcliffe et al., 2006, 2009; DeLong et al., 2007). Social integration has been theorized by Tinto (1975, 1993) to be a key contributor to student persistence, and ''social fit'' indicators as measured by living learning communities have shown to play a key role in our understanding of the success of students (Matthews 1996; Tinto 1998). Ishitani (2006) found that on average, low income students are more likely to drop out early and less likely to graduate than students from higher income families. These previous studies illustrate the complexity of the underlying process that influences a student's choice to persist in higher education.

One study conducted by Jones-White et al. (2009) even furthered the investigation on student retention and graduation by redefining student success. The researchers strongly argued that the depiction of student success have been compromised by the existing definitions of retention and graduation rates. They believe that the pool of data should not be limited to completion of degree at the institution of entry only but should also be extended even to transfer institutions. However, to accommodate such extended definition of multi-institutional student success, more sophisticated modelling techniques were utilized like multinomial regression method.

Influence of Academic Factors and Student Success in Physics

A number of studies have already been conducted to specifically study the correlation of pre-college preparations as reflected in the student's high school ratings and standardized test scores, and their success in college physics courses. Previous studies carried out at individual colleges tested the association between variables such as high school grades, coursework, or standardized test scores and success in college physics.

Gifford and Harpole (1986) found that for 248 Mississippi State University students, high school math grades, taking high school physics, and high amounts of laboratory time were associated with high grades in college physics. Hart and Cottle (1993) found that high performing introductory college physics students at Florida State University had performed well in high school math and had taken high school physics. Alters (1995) replicated this study at the University of Southern California with 161 students finding similar results. Each of these studies examined only a few variables and relied upon simple correlations with college grades. One finding was that those students who do well in college science often did well in high school mathematics.

They say that “mathematics is the language of physics”. In the physics department, it has been a perennial observation that certain students really do work hard in their study of physics and have not given up trying, yet still failed to master the subject. This kind of scenario suggests that there are certain intellectual barriers that keep these students from successfully thriving in the program and halts their learning process. Among the many possible factors that top the list would be the student’s mathematical prowess or the least a mastery of the basic mathematical skills.

One study stated that one of its most notable results is that students’ math ability had a significant impact on the student’s outcome measures, indicating that a certain level of mathematical understanding is required for performing well at physics examinations. (Korpershoek, 2014). This is also supported by Hudson and McIntire’s (1977) observation that for many students in their study mathematical ability correlates with performance in introductory physics course. It has also been found that mathematical skills together with the ability to employ formal operational reasoning are one of the several factors necessary for success in physics courses (Liberman and Hudson, 1979).

As far as foreign languages are considered, English is the language which has highest influence on Philippine Educational System. This was due to the fact that the Philippines had been for a while a territory of the United States of America. This choice of English as a common language was also dictated by the fact that English has been accepted as a language of communication almost all over the world. No wonder, English has been considered as a second language in school education.

For higher education English language is the commonly used language as a medium of instruction most especially for science and applied science programs. The BS Physics curriculum is no exception to this reality as English has been recognized as the official language and medium of instruction at university level. Most of the available study material and literature in physics is in English. Due to this reality, physics students are required and expected to have a certain level of comprehension ability when it comes to reading text books and related literature; and a good command of the language in written and verbal communication.

Students with a high score on reading ability benefit from this skill when taking their examinations in advanced mathematics, physics, and chemistry according to Korpershoek (2014). Making sense of mathematical word problems evidently requires a certain level of text comprehension

(Verschaffel et al., 2000). To construct meaning from the story or real-life scenario presented in mathematical word problems, students obviously need a certain level of reading ability (Adams, 2003; Graesser et al., 1997).

Korpershoek (2014) reiterated that it may be justifiable to argue that reading comprehension should be an integrated part of mathematics, physics, and chemistry courses. Therefore, further examination of this relationship is certainly needed to increase our understanding of the significance of reading ability for educational attainment in tertiary education.

Based on the collected literatures, all the identified researches sought to determine the correlation of student success to varied cognitive and non-cognitive factors. Many of the researches are in the field of education, and some allied sciences. Although there were literatures specifically tackling on physics, all of the ones cited are in relation to student success in Introductory Physics as a course and not as a degree program.

The variables enumerated by the researchers can be grouped into two (2) general areas of cognitive and non-cognitive variables. Cognitive variables included standardized test scores, grade point averages, etc. Non-cognitive variables included the student background/demographics, social integration and the likes.

The review of the literature reveals that the predictors of success are the academic performance such as grade point average and standardized test scores and individual demographics. However, in most of the studies conducted limited citations can be done on the correlation of early college ratings of the entrants with their academic outcomes in the degree program they are enrolled in. The researcher of this study would like to solely use the early academic standing of the freshmen Physics students in the area of Mathematics, Physics and English/Language and their Pre-college preparation as reflected in the CET Results to predict the outcomes of the student in the degree program.

Statement of the Problem

This study aimed to generate a mathematical model that can be utilized as an early detection tool for Physics Students study outcome (successful or unsuccessful) using only the student's first year academic standing (English, Math and Science) and College Entrance Test (CET) Results.

More specifically this paper aimed to identify which cognitive variable is the best predictor for a student's study outcome in the Physics program and to validate the developed model in determining the study outcome of Physics students studying in the university.

Methodology

Research Design

The study employed the descriptive correlation method research design to determine the association of the independent variables which is the academic variable to include the College Entrance Test, Physics, Algebra, Trigonometry, Calculus with Analytic Geometry, and English, with the dependent variable which is the students' study outcome. The independent variables were analyzed with the dependent variables, this time using Logistic Regression statistics for the formulation of the predictor model.

Sampling and Population

The study used purposive sampling considering the limited number of students in the Physics program. There was a total of one hundred twelve (112) BS Physics students in the given time-range. The distribution of samples derived per year was based on the actual number of enrollees from Ten (10) consecutive academic years. The sources of data for the study included the transcript of records (TOR) and College Entrance Test (CET) results of the students.

Data Analyses Procedures

The Point-biserial correlation analysis was utilized in the ratings for the five (5) subjects with the CET Results and the success response of the students to determine whether significant correlation exists between the independent and dependent variables. Logistic Regression was also used to determine whether academic variables to include the CET Results and the ratings on the five (5) subjects could predict success in the BS Physics Program and a predictive model was generated thereafter. The Model was also validated through Hosmer-Leme show 'Goodness-of-fit' test.

Results and Discussions

This chapter presents the data gathered and the results obtained from the study. The results are presented, analyzed, and interpreted based on the specific problems as presented above. The table below shows the frequency and percentage distribution of grades of the BS Physics students in the different subjects.

Table 1: Frequency and Percentage Distribution of Students' Grades in the Different Subjects*

Ratings	Subjects									
	Algebra		Trigonometry		Calculus		Physics		English	
	f	%	f	%	f	%	F	%	f	%
1.00	5	4.5	2	1.8	1	0.9	1	0.9	0	0
1.25	8	7.1	7	6.3	2	1.8	1	0.9	6	5.4
1.50	13	11.6	8	7.1	6	5.4	3	2.7	11	9.8
1.75	3	2.7	8	7.1	8	7.1	10	8.9	13	11.6
2.00	5	4.5	6	5.4	6	5.4	8	7.1	13	11.6

2.25	4	3.6	7	6.3	5	4.5	13	11.6	18	16.1
2.50	13	11.6	8	7.1	5	4.5	17	15.2	18	16.1
2.75	12	10.7	9	8.0	4	3.6	9	8.0	10	8.9
3.00	22	19.6	34	30.4	25	22.3	17	15.2	4	3.6
5.00	27	24.1	23	20.5	50	44.6	33	29.5	19	17.0
Total	112	100	112	100	112	100	112	100	112	100

*Subject grades are coded following the university grading system 1 to 5. (1 serves as the highest possible grade)

Table 1 shows that majority of the physics students (54.4% and 52.7%) obtained a grade from 2.75-5.0 in Algebra and Physics respectively. In the cases of Trigonometry and Calculus, most of the students (at respective percentages of 50.9% and 66.9%) got ratings from 3.0-5.0. In the context of the university grading system, these ratings are in the descriptive marks of Passing for 3.0 and Failed for 5.0. Across all the five (5) subjects, Calculus with Analytic Geometry seemed to be the most difficult for the respondents.

For English however, most of the students (at 55.4%) got a grade from 1.75-2.5. Descriptively, these ratings are from good to very good. This information indicates that the BS Physics first-year students may have had a more challenging time in their major subjects as compared to English 101.

Table 2 presents the frequency and percent distribution of the ratings on the College Entrance Test (CET) of the BS Physics graduates.

Table 2: Frequency and Percent Distribution of Overall Percentile Rank (OAPR) of the Physics Graduates in the College Entrance Test*

Range of Percentile Rank in CET	Frequency	Percent
100-91	38	33.9
90-81	21	18.8
80-71	19	17.0
70-61	12	10.7
60-51	13	11.6
50-41	5	4.5
40-31	3	2.7
30-21	0	0
20-11	1	0.9
Total	112	100

*CET ratings are coded in percentile rank.

The ratings of the graduates on their College Entrance Test indicate that a majority of them had a college entrance examination percentile rank score between 81-100 comprising 52.7% of the sample. These scores indicate that more than half of the respondents belonged to the top 20% of those who took the entrance test during their time.

Correlational Analysis of the Variables

Point-biserial correlation analysis was conducted to examine the relationship between the dependent variable, student study outcome, and the various potential predictors that are found to be contributory to the success of the students in the BS Physics Program. Table 3 displays the predictors/factors with their correlation values.

Table 3: Correlations of Predictor Variables to Student Study Outcome (N=112).

No.	Predictor Variables	Pearson r	p-value	Description
1	Algebra	- 0.581**	0.00	Moderate Negative Correlation
2	Trigonometry	- 0.563**	0.00	Moderate Negative Correlation
3	Calculus	- 0.728**	0.00	High Negative Correlation
4	Physics	- 0.613**	0.00	Moderate Negative Correlation
5	English	- 0.511**	0.00	Moderate Negative Correlation
6	CET (Percentile)	0.397**	0.00	Low Positive Correlation

** Correlation is significant at the 0.01 level (2-tailed)

As can be observed, Algebra, Trigonometry, Calculus, Physics and English results scores are negatively and significantly correlated with the Student Study Outcome, indicating that students with higher scores on these scales (Grading System; 1.0 as the highest) have lower probability of success in the BS Physics program; however, lower scores would mean high probability of success in the program. Furthermore, CET is positively and significantly correlated with the dependent variable, indicating that those who belong to higher percentile ranks tend to have more probability of success. This coincides with the conclusions of the study done by Daempfle (2003) in which the cognitive variables were accounted to have a positive correlation toward success in college work.

The correlation found between student success and each of the factors indicated conformity with some studies. The research conducted by Rohr (2009) showed that Scholastic Aptitude Test and College Preparatory GPA were significant predictors of retention of science, mathematics, engineering technology and business students. The findings of Gilmer (2007) reported that success in Mathematics or a good performance in quantitative skills is a key to success in science, technology, engineering and mathematics majors. The study of Rubio (2011) found that the students who are good in English and Filipino and who speak the two languages are also good in science.

Logistic Regression Analysis and Prediction Model

To generate the desired prediction model for the study outcome of students, Binary Logistic Regression analysis was applied. The model created from the six predictor variables produced a R^2 (Cox & Snell) =0.496 and R^2 (Nagelkerke) =0.663 and an Omnibus Tests significance value of p= 0.000 as shown in Table 4 and 5.

Table 4: Omnibus Tests of Model Coefficients for Logistic Regression

		Chi-square	Df	Sig.
Step 1	Step	76.826	6	.000
	Block	76.826	6	.000
	Model	76.826	6	.000

The Omnibus Tests of Model Coefficients reports the chi-square associated with each step in a stepwise model. Here, there is only one step from the constant model to the block containing predictors so all three values are the same. The significance value or p-value indicates our model (the generated model) is significantly different from the constant only model (null model); meaning there is a significant effect for the combined predictors on the outcome variable. A significance value of less than 0.05 which is indicated in the above data suggests that the generated model is a good one.

Table 5: Logistic Regression Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	77.867[a]	0.496	0.663

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

The Model Summary table displays the -2 LL which is a likelihood ratio and represents the unexplained variance in the outcome variable, therefore, the smaller the value, the better the fit. As shown in Table 5, the -2 LL has a value of 77.867 and is substantially lower than the one given by the null model which is 154.693.

The Nagelkerke estimate is a fairly good way of evaluating a model fit. Since it is calculated in such a way as to be constrained between 0 and 1; a value of 0.663 indicates a fairly good model. In the same way, the larger the Cox & Snell estimate is the better the model; but it can be greater than 1.

Table 6 : Variables in the Equation for the Logistic Regression Model

Step 1[a]	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)

								Lower	Upper
	Algebra	-.002	.434	.000	1	.996	.998	.426	2.334
	Trigo	-.214	.405	.279	1	.597	.808	.365	1.785
	Calculus	-1.140	.431	6.992	1	.008	.320	.137	.745
	Physics	-.136	.437	.096	1	.756	.873	.370	2.057
	English	-1.243	.666	3.484	1	.062	.288	.078	1.064
	CET	-.005	.021	.059	1	.808	.995	.954	1.037
	Constant	8.117	3.155	6.617	1	.010	3351.859		

a. Variable(s) entered on step 1: Algebra, Trigo, Calculus, Physics, English, CET r.

Thus, the generated statistical model from the Logistic Regression is in the form as follows:

$$\text{Probability of Success} = \frac{1}{1 + e^{-(7.9-0.002*\text{Algebra}-0.214*\text{Trigo}-1.14*\text{Calculus}-0.136*\text{Physics}-1.243*\text{English}-0.005*\text{CET})}}$$

However, by looking at Table 7, we see that our model fails the Hosmer and Leme show Test of ‘Goodness-of-Fit’. As with most chi-square based tests, it is also prone to inflation as sample size increases. Here, we see model fit is unacceptable χ^2 (9) = 18.709, p = .016, which indicates our model predicts values that are significantly different from what we observed. As a rule-of-thumb, we want the p-value to be greater than the established cut-off (generally 0.05) to indicate good fit.

Table 7: Hosmer and Leme show Test

Step	Chi-square	df	Sig.
1	18.709	8	.016

Generating a Reduced Model

Due to the unacceptable fit of the produced model, there is a need to generate a reduced one with lesser independent variables as predictors. Referring to Table 6, it is observed that Algebra and CET are the least likely to predict the success of the BS Physics students in the program as depicted in their β coefficients (-0.002 and -0.005 respectively). It is deduced then that these two variables will be omitted as outcome predictors from the model generation.

The new model created from the four predictor variables produced a R2 (Cox & Snell) =0.496 and R2 (Nagelkerke) =0.663 and an Omnibus Tests significance value of p= 0.000 as shown in Table 8 and 9.

Table 8: Omnibus Tests of Model Coefficients for Logistic Regression

		Chi-square	Df	Sig.
Step 1	Step	76.765	4	.000

	Block	76.765	4	.000
	Model	76.765	4	.000

A significance value of less than 0.05 in the Omnibus Tests of Model Coefficients which is indicated in the above data suggests that the generated model is a good one.

Table 9: Logistic Regression Model Summary (Reduced)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	77.928a	0.496	0.663

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

The Model Summary table displays that the -2 LL has a value of 77.928 and is substantially lower than the one given by the null model which is 154.693. This indicates a measure of a good fit. Likewise, a value of 0.663 in the Nagelkerke estimate indicates a fairly good model.

Table 10 gives the new set of variables for the reduced Logistic Regression Model. This new model will be subjected to Hosmer and Leme show Test of 'Goodness-of-Fit' for validation and test of prediction accuracy.

Table 10 : Variables in the Equation for the Logistic Regression Model (Reduced)

Step 1a		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
	Trigo	-.186	.380	.239	1	.625	.830	.394	1.750
	Calculus	-1.135	.375	9.169	1	.002	.321	.154	.670
	Physics	-.130	.431	.091	1	.763	.878	.878	2.043
	English	-1.190	.616	3.724	1	.054	.304	.304	1.019
	Constant	7.477	1.665	20.157	1	.000	1766.982		

a. Variable(s) entered on step 1: Trigo, Calculus, Physics, English

Hence, the generated reduced statistical model from the Logistic Regression is as follows:

$$\text{Probability of Success} = \frac{1}{1 + e^{-(7.477 - 0.186 * \text{Trigo} - 1.135 * \text{Calculus} - 0.13 * \text{Physics} - 1.190 * \text{English})}}$$

The Variables in the Equation table (Table 10), shows the logistic coefficient (B) for each predictor variable. The logistic coefficient is the expected amount of change in the student study outcome for each one unit change in the predictor. The closer a logistic coefficient is to zero, the less influence it has in predicting the outcome.

That being the case, we can say that the coefficient of Trigonometry can be interpreted as for every 1 grade value increase in Trigonometry the odds of having a successful outcome is by exp(-0.209) = 0.811 times. Hence, if a previous chance to success is 0.9, after 1 grade value increase in Math 103, chance of success is now 0.9*0.811 = 0.73. Since, the coefficient values show which variables would be good predictors of the outcome, we can say that Calculus has the biggest influence on predicting the success of students in the BS Physics program.

Logistic Regression Model Validation

The Hosmer and Leme show Test table (Table 11) is the preferred test of goodness-of-fit of our prediction model. With the new model, we can observe that the model fit is acceptable χ^2 (9) =15.14, p = 0.056, which indicates our model predicts values not significantly different from what we observed.

Table 11: Hosmer and Leme show Test (Reduced)

Step	Chi-square	df	Sig.
1	15.7140	8	**.056**

Table 12 gives the Classification Table for Prediction Model. The Classification Table shows how well our full model correctly classifies cases. A perfect model would show only values in the diagonal--correctly classifying all cases. For this model, the overall percentage is 84.8% accurate; which is good indicator of predictability.

Table 12: Classification Table for Prediction Model[a]

	Observed		Predicted		
			Outcome		Percentage Correct
			Failure	Success	
Step 1	Outcome	Failure	48	12	80.0
		Success	5	47	90.4
	Overall Percentage				84.8

a. The cut value is .500

Conclusion and Recommendations

Point-biserial correlation was used to determine the extent of the relationship of the six (6) cognitive variables to the study outcome of the students (Successful or Failure). Binary Logistic Regression analyses were also ran to generate a statistical model in predicting the outcome of the students in the BS Physics Program and to determine which among the academic variables significantly influence these predictions.

The following are the results of the study:

1. Point-Biserial Correlation at $p<0.05$ disclosed that all of the cognitive variables (Algebra, Trigonometry, Calculus, Physics, English and CET Results) were found to be moderately correlated with the student study outcome.
2. Binary Logistic Regression analyses generated the prediction model in this form:

$$\text{Probability of Success} = \frac{1}{1 + e^{-(7.477 - 0.186*\text{Trigo} - 1.135*\text{Calculus} - 0.13*\text{Physics} - 1.190*\text{English})}}$$

3. Binary Logistic Regression analyses specified that the grades in Calculus and English have the most significance in predicting the success of the Students in the BS Physics program, followed by Trigonometry and Physics. CET Results and Algebra were found to have no influence in predicting the study outcomes of the students.
4. The generated statistical model is found to be valid and acceptable with overall percentage accuracy of 84.8%.

In the emerging field of educational data mining, a strong bias towards data-rich digital learning environments is the current state of affairs. In this study, the situation at the Physics department of a particular university was considered. The department claims to be able to properly distinguish the potentially successful students from amongst the 1st year influx before the end of the academic year, but the selection is only loosely based on assumed student similarities over the years. There was no thorough analysis. The generated model now can be applied in properly assessing the future academic standing of First Year BS Physics students in the program in relation to their CET results and academic performance. The prediction can lead to appropriate interventions, the nature of which may be drawn from the students' perceived causes for attrition and can then be implemented to help those students found to have a high probability of being unsuccessful in the program.

The results of the study show that the freshmen's subjects' academic performance can be considered as strong predictors of success. By determining the strength of the predictor's quality, the researcher can now focus in determining the necessary information to make the current assessment work. If the predictor quality is high, the researcher's interests can be directed towards: (1) using the predictor as a back-up of the current assessment process; (2) identifying success-factors specific to the BS Physics program; (3) identifying what data might result in a further increase of the predictor quality, and as a consequence, collect these data; (4) modifying the assessment process time-line, resulting in an earlier prediction, ideally even before entering the study. Furthermore, the knowledge of the strength of the predictors for academic success can be utilized as a means to gain understanding of success and risk factors regarding the curriculum. Awareness of these factors by teachers, education personnel and management will help to select appropriate measures to support the risk group, eventually resulting in a decrease of the drop-out rate.

Based on the findings, the researcher recommends the following:

1. To conduct a similar study, using not only the CET General Results but the subtest component scores too.
2. To conduct intervention programs that address psychosocial factors like motivation, commitment, time management, and career orientation and opportunities and the same factors be considered in future studies.
3. To include non-academic variables as well in a similar study.

References

Aina, J. K., Ogundele, A. G., & Olanipekun, S. S. . (2013). Students' Proficiency in English Language Relationship with Academic Performance in Science and Technical Education. American Journal of Educational Research, 1(9), 355-358 .

Aljohani, O. (2016). A Comprehensive Review of the Major Studies and Theoretical Models of Student Retention in Higher Education. Higher Education Studies, 6(2), 1−18. https://doi.org/10.5539/hes.v6n2p1 .

Arboretti Giancristofaro, R., & Salmaso, L. (2003). Model Performance Analysis and Model Validation in Logistic Regression. Statistica, 63(2), 375-396.

Awofala, A. A., Awofala, A. A., Nneji, L. M., & Fatade, A. O. (2012). The Relation between Language and Science, Technology, and Mathematics in Nigerian Schools. International Journal of Mathematics Trends and Technology, 3(3), 118-124 .

Derrick, M.E., & Derrick, F.W. (2002). Predictors of success in physical chemistry. Journal of Chemical Education, 79 (8) , 1013-1016.

Fagan, D. . (1997). Reading in Advanced level physics. Physics Education, 32(6), 383.

Howard, B. & Sharpe, L. (2019). The Summer Bridge Program: An Effective Agent in College Students'Retention. Journal of Interdisciplinary Studies in Education, 7(2), 20-30.

Hudson, H.T.,& McIntire, W.R. . (1977). Correlation Between Mathematical Skills and Success in Physics. American Journal of Physics, 45(5), 470-471.

Korpershoek, H., Kuyper, H., & van der Werf, G. (2015). The Relation Between Students' Math and Reading Ability and Their Mathematics, Physics, and Chemistry Examination Grades in Secondary Education. International Journal of Science and Mathematics Education, 13(5), 1013-1037 .

Kuncel, N. R., & Hezlett, S. A. (2007). ASSESSMENT: Standardized Tests Predict Graduate Students' Success. Science, 315(5815), 1080–1081. doi:10.1126/science.1136618 .

Liberman, D., & Hudson, H.T. . (1979). Correlation Between Logical Abilities and Success in Physics. American Journal of Physics , 47(9), 784-786.

Orleans, A. (2007). The Condition of Secondary School Physics Education in the Philippines: Recent Developments and Remaining Challenges for Substantive Improvements. The Australian Educational Researcher, 34(1), 33-54.

Pingry O'Neill, L.N.; Markward, M.J.; French, J.P. . (2012). Predictors of Graduation Among College Students with Disabilities . Journal of Postsecondary Education and Disability, 25(1), 21 - 36.

Rana, S., Midi, H., & Sarkar, S.K. (2010). Validation and Performance Analysis of Binary Logistic Regression Model. Proceedings of the WSEAS International Conference on ENVIRONMENT, MEDICINE and HEALTH SCIENCES, (pp. 51-55).

Sezgin Selcuk, G. (2010). Correlation study of physics achievement, learning strategy, attitude and gender in an introductory physics course. Asia-Pacific Forum on Science Learning and Teaching, 11(2), 4. https://www.researchgate.net/publication/50257362_Correlation_study_of_physics_achievement_learning_strategy_attitude_and_gender_in_an_introductory_physics_course.

Tai, R., Sadler, P., & Mintzes, J. (2006). Factors Influencing College Science Success. Journal of College Science Teaching, 56-60.

Thomas, E. W., Marr, M. J., Thomas, A., Hume, R. M., & Walker, N. . (n.d.). Using Discriminant Analysis to Identify Students at Risk. Technology-Based Re-Engineering Engineering Education Proceedings of Frontiers in Education FIE'96 26th Annual Conference, doi:10.1109/fie.1996.569940 .

Tinto, V. (1999). Rethinking the First Year of College. NACADA Journal, 1-7. https://www.researchgate.net/publication/228860241_Taking_Student_Retention_Seriously_Rethinking_the_First_Year_of_College/link/571d595c08ae6eb94d0e512b/download.

Witz, K., Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1990). Applied Statistics for the Behavioral Sciences. Journal of Educational Statistics, 15(1), 84. doi:10.2307/1164825 .

Worley, C. L. (2007). At-risk students and academic achievement : The relationship between certain selected factors and academic success. (Unpublished Dissertation). Virginia Polytechnic and State University. https://www.researchgate.net/publication/35875377_At-Risk_Students_and_Academic_Achievement_The_Relationship_Between_Certain_Selected_Factors_and_Academic_Success.