2018 ASEE Annual Conference & Exposition
Paper ID #21226

A Retention Model for Community College STEM Students

Jennifer Snyder, Valencia College

Jennifer Snyder is a Ph.D. candidate in the Department of Engineering Management and Systems Engineering at Missouri University of Science and Technology. She received her B.S. and M.S. in Chemistry from Missouri State University in Springfield, Missouri. She is a dean of science for Valencia College in Orlando, Florida.

Dr. Elizabeth A. Cudney, Missouri University of Science & Technology

Dr. Elizabeth Cudney is an Associate Professor in the Engineering Management and Systems Engineering Department at Missouri University of Science and Technology. She received her B.S. in Industrial Engineering from North Carolina State University, Master of Engineering in Mechanical Engineering and Master of Business Administration from the University of Hartford, and her doctorate in Engineering Management from the University of Missouri – Rolla. In 2017, Dr. Cudney received the 2017 Yoshio Kondo Academic Research Prize from the International Academy for Quality for sustained performance in exceptional published works. In 2014, Dr. Cudney was elected as an ASEM Fellow. In 2013, Dr. Cudney was elected as an ASQ Fellow. In 2010, Dr. Cudney was inducted into the International Academy for Quality. She received the 2008 ASQ A.V. Feigenbaum Medal and the 2006 SME Outstanding Young Manufacturing Engineering Award. She has published six books and over 65 journal papers. She holds eight ASQ certifications, which include ASQ Certified Quality Engineer, Manager of Quality/Operational Excellence, and Certified Six Sigma Black Belt, amongst others. She is a member of the ASEE, ASQ, IISE, and the Japan Quality Engineering Society (JQES).

A Retention Model for Community College STEM Students

Abstract
The number of students attending community colleges that take advantage of transfer pathways to universities continues to rise. Therefore, there is a need to engage in academic research on these students and their attrition in order to identify areas to improve retention. Community colleges have a very diverse population and provide entry into science, technology, engineering, and math (STEM) programs, regardless of student high school preparedness. It is essential for these students to successfully transfer to universities and finish their STEM degrees to meet the global workforce demands. This research develops a predictive model for community college students for degree completion using the Mahalanobis Taguchi System and regression. Data collected from a Midwest community college over a five-year period in three specific associate degree programs will be used for the study. The study identified 92 students that completed a STEM degree within three years, while 730 students were not able to complete the degree within that period or at all. The research illuminates specific areas of concern related to community college students and better informs transfer institutions about this important sector of transfer students. Especially revealing is the important predictive factors traditionally found in research for STEM retention had very low correlation for this set of community college students. This research reinforces the need to investigate community college students more closely and through a different lens.

Keywords: Predictive Analytics, Community College, Education, Mahalanobis Taguchi System, Diversity

Introduction

Community colleges play a pivotal role in higher education. One area of growth has been in the area of serving as a pipeline to transfer universities (Adelman, 2005). These are students that begin their higher education path at a community college; either completing an associate's degree or transferring after taking some general education courses. Many universities find themselves in a position where their growth is dependent upon transfer students. This process will continue to expand due to the confluence of rising tuition, student need for remediation, rise in technical degrees, and desire to have a greater percent of citizens obtaining a post-secondary credential (Cohen, Kisker, & Brawer, 2014).

One of the most critical student populations are those pursuing a science, technology, engineering, or math (STEM) degree (Hoffman, Starobin, Laanan, & Rivera, 2010). It has been reported that roughly 50% of graduates with bachelor degrees in STEM fields took some courses at a community college. Chen (2013) reports that community college students declaring STEM degrees have a higher attrition rate (69%) compared to university students (48%). The report further found that of the community college students that left STEM half changed majors, while the other half left the system without a degree or certificate. As the interest in community colleges has grown, the research interest has been slow to catch up (Starobin & Laanan, 2010). The causes of attrition from STEM degrees is not well researched and reported for this sector of students. A majority of STEM retention models and studies deal with data collected from traditional university students. The factors available for investigation are limited and might not be available or indicative for community college students (Snyder & Cudney, 2017). There is a dearth of research into community college STEM students and their particular risk factors that would prevent them from completing a STEM degree within 150% time to degree, which is three years.

This research seeks to answer some of the questions surrounding this population of students. The research uses data collected from a community college in the Midwest. The Mahalanobis Taguchi System (MTS) is used for pattern recognition and a predictive model is developed using logistic regression. The following questions are investigated:

1. Can the Mahalanobis Taguchi System forecast important variables used for a STEM retention prediction model?
2. Do community college students have substantially different risk factors than traditional university students?

The remainder of this paper is structured into the following sections: literature review and background on community colleges, data analysis and predictive model development, validation, and comparison to university models.

Literature Review

Community colleges were born out of a need for higher education and technical training (Cohen et al., 2014). Joliet Junior College, founded in 1901, was the first public community college. The primary mission of community colleges has not changed greatly, but there has been refinement through the years to serve the changing population and economy (Cohen et al., 2014; Hoffman et al., 2010). Community colleges are more agile and responsive to market demands on a local level, which can be seen by evaluating the technical degree landscape.

Community college students are reflective of the region in which the college is located due to most community colleges being commuter campuses. Further, a greater number of minority and lower socioeconomic students (SES) attend community colleges (Costello, 2012; Horn & Nevill, 2006). Carnevale and Strohl (2010) report that bottom quartile SES outnumber top quartile SES by almost 2 to 1 at community colleges, while top quartile SES outnumber bottom quartile SES at competitive colleges by almost 10 to1. Community college students are more likely to attend college part time and work full time (Horn & Nevill, 2006). Costello (2012) reported that twice as many students at community colleges are parents compared to universities. Community colleges are usually open access; therefore, there are no entrance requirements such as standardized exam score benchmarks. In fact, it is estimated that more than 60% of community college students receive some remedial education upon entrance to college (Crisp & Delgado, 2014). These factors contribute to the outcomes experienced at community colleges.

As the twenty first century moves forward, the country has been charged with increasing the number of STEM graduates to meet the growing global demands (Committee on Science, 2007). In 2012, The President's Council of Advisors on Science and Technology (PCAST) produced a report outlining steps necessary to reach the goal of increasing STEM graduates by one million (Olson & Riordan, 2012). This goal is only surmountable if retention rates are increased. It has been reported that a ten percent increase in retention rates will garner three-quarters of the goal (Carver et al., 2017; Graham, Frederick, Byars-Winston, Hunter, & Handelsman, 2013). If retention is not impacted, then the number of students declaring a STEM degree must increase. Student interest in STEM has remained unchanged for years (Hurtado, Eagan, & Chang, 2010). The only area of increase of students declaring STEM degrees is in the Hispanic and African American population. Numbers show that for the first time the declaration rates are equal for all students (Hurtado et al., 2010).

This increase in minorities and underrepresented populations declaring STEM degrees is needed to diversify the workforce (Terenzini, Lattuca, Ro, & Knight, 2014). There has been a call for diversification for years. The needle has moved on intent, but the retention and completion rates are slow to move (Hurtado et al., 2010).

The transfer pathway should be a critical component in this effort. There should be more done to increase the retention and completion of community college STEM students. This is the importance of investigating a predictive model built with community college student data. If these students are demographically different, then the predictive models and risk factors are likely very different.

Data Analysis and Prediction Model

Data for this research was collected from a community college in the Midwest. This community college is ideal for data collection, because it has associate degrees in STEM fields that students can declare from the beginning. The raw data, collected over a five-year period, identified 177 students that completed an associate's degree in either chemistry, biology, or engineering; while 727 students were not successful. The unsuccessful students either withdrew from the college or switched degree to non-STEM fields.

The raw data illuminates one of the problems associated with an open access institution such as a community college. There is considerable missing data, inaccurately reported data, and many students did not have standardized exam scores. The descriptive statistics for the raw data are shown in Table 1.

The Mahalanobis Taguchi System was chosen for the process of identifying important variables. MTS is a pattern recognition method used in various industries (Ghasemi, Aaghaie, & Cudney, 2015). Ghasemi et al. (2015) reviews the approach of MTS, which involves dividing the data into normal and abnormal groups. Woodall et al. (2003) breaks MTS into four steps or stages:

Stage 1: The variables are identified that will be defined as normal and abnormal. For this research, the completion of a STEM associate's degree within three years is normal and not completing the STEM degree within three years is abnormal. The normal data is standardized and a Mahalanobis space is determined using the normal data, which is referred to as the reference space.

Stage 2: The abnormal items, test group, are selected and the Mahalanobis distance (MD) is calculated. In this research, the MD for the abnormal group is 6.0863. This MD value indicates that the scale is appropriate as the MD for the abnormal group is higher than the MD for the normal group, which was verified with this data.

Stage 3: In this stage, the orthogonal arrays (OA) and signal-to-noise (S/N) ratios are calculated and used to determine the most useful set of predictive variables. Larger S/N ratios are preferred and indicate a possible useful predictive variable.

Stage 4: The variables that were identified as significant due to a positive S/N are used to develop a forecasting model.

Table 1. Descriptive Statistics of Raw Data

Completers					
Factor	N	Mean	Median	Range	
				Minimum	Maximum
Age	177	25.62	22.00	18.00	53.00
ACT Comp	106	23.43	23.00	12.00	34.00
ACT Eng	107	22.72	23.00	13.00	34.00
ACT Math	107	23.85	24.00	13.00	35.00
ACT Read	106	23.96	24.00	13.00	36.00
High School GPA	145	4.49	3.67	1.17	86.53
College GPA	177	3.31	3.36	2.00	4.00
Non-completers					
Factor	N	Mean	Median	Range	
				Minimum	Maximum
Age	727	24.07	21.00	16.00	65.00
ACT Comp	322	21.85	22.00	11.00	33.00
ACT Eng	329	21.30	21.00	7.00	35.00
ACT Math	329	21.82	22.00	13.00	33.00
ACT Read	329	22.30	22.00	9.00	36.00
High School GPA	460	3.77	3.35	1.00	91.38
College GPA	637	2.24	2.54	0.00	4.93

As an open-admission institution, data such as high school GPA and ACT scores are not required; therefore, many students had incomplete records. The students that did not report test scores or high school information were removed from the sample. The final data set had 97 successful (normal) students and 32 unsuccessful (abnormal) samples. The results are summarized in Table 2.

A larger S/N indicates a strong significance for that factor, which implies that part-time student status and college GPA are the most important factors to explore. It is interesting to note that ACT math did not have a large S/N ratio, which contrasts with most STEM retention models that usually weight math scores heavily. MTS results indicate that the factors with positive S/N ratios are important for forecasting the completion of a STEM degree for community college students. It is not surprising that part-time status has a significant impact considering the three-year completion window. Students that attend school part-time find it very difficult to complete a rigorous degree in three years. This is an important factor to consider when advising students.

For the development of the predictive algorithm, logistic regression was performed using stepwise selection of the terms above. The limit to enter and remove variables in the model (alpha, α) was set to 0.05. The results of the regression are shown in Table 3, which indicates gender, college GPA, and enrollment status are significant variables for prediction.

Table 2. Results of the Mahalanobis Taguchi System

Factor	S/N ratio	Include in model?
Part-time student status	6.2493	Yes
College GPA	1.4484	Yes
ACT comprehensive	0.5788	Yes
Degree declared (biology, chemistry, engineering)	0.4614	Yes
Gender	0.4381	Yes
ACT Math	0.3211	Yes
ACT Reading	0.1205	Yes
Plan to work while attending college	-0.1104	No
ACT English	-0.1493	No
Age	-0.3031	No
High school GPA	-0.3179	No
First-Generation college student	-0.6895	No

Table 3. Stepwise Selection of Terms

Deviance Table

Source	DF	Adj Dev	Adj Mean	P-Value
Regression	3	120.825	40.2752	0.000
College GPA	1	35.717	35.7174	0.000
PT	1	84.352	84.3517	0.000
Gender	1	4.740	4.7395	0.029
Error	125	23.705	0.1896	
Total	128	144.531		

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC
83.60%	81.52%	31.71

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
College GPA	23.5598	(3.9198, 141.6068)

Odds Ratios for Categorical Predictors

Level A	Level B	Odds Ratio	95% CI
PT			
1	0	0.0002	(0.0000, 0.0100)
Gender			
1	0	18.4118	(0.8762, 386.8742)

Odds ratio for level A relative to level B

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

PT	**gender**				**Student Profile**
0	0	Y'	=	-5.926 + 3.160 College GPA	Full-time/Female
0	1	Y'	=	-3.013 + 3.160 College GPA	Full-time/Male
1	0	Y'	=	-14.61 + 3.160 College GPA	Part-time/Female

1	1	Y'	=	-11.69 + 3.160 College GPA	Part-time/Male

Validation and Comparison

Overall model evaluation is determined by whether the model is better than the intercept-only model. If the values of the coefficients for the variables in the equation are zero, then the model is not an improvement on the intercept-only model. Figure 1 and Table 4 indicate the model is better at predicting the probability of completion, with it predicting 98% correctly for the successful completion and 91% for non-completion.

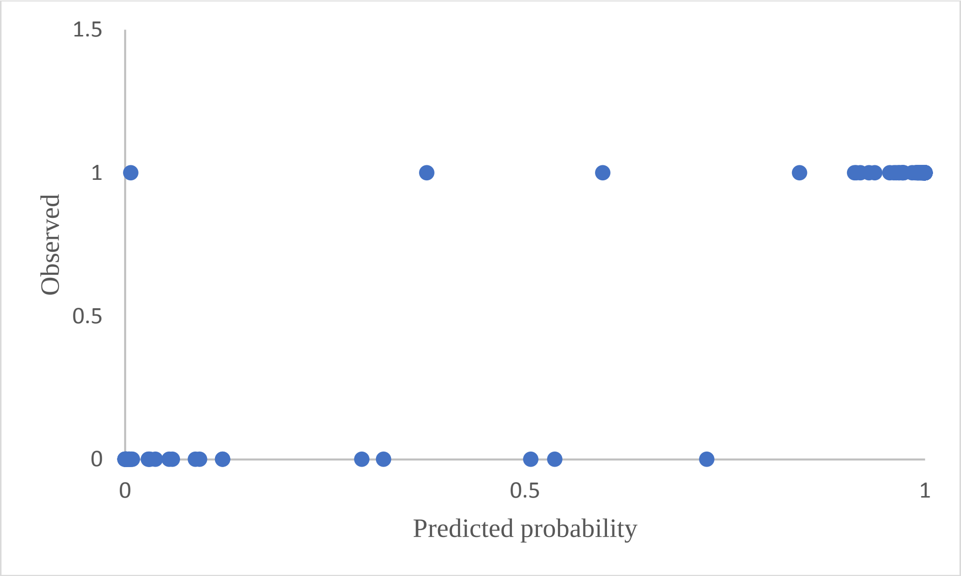

Figure 1. Observed values compared to the predicted probabilities

Table 4. Correlation of predicted and actual

		Predicted	
		Yes	No
Actual	Yes	95	2
	No	3	29

The adjusted R^2 of the model indicates 81.52% of the variation in the completion rates of a STEM degree for community college student can be predicted by the model, which includes demographic and enrollment data only.

From an advising perspective, this is a powerful model if the student has a college GPA. The goal is to predict success and advise the student accordingly. Recognizing the importance of GPA on completion, a regression analysis was performed to predict college GPA for community college students. Stepwise regression was performed on the data using an α of 0.05. The results are provided in Table 5.

Performing this regression was useful to understand the factors that could predict college GPA, which is a strong predictor of completion. The interesting significant predictor is ACT reading scores. Currently, many community colleges are eliminating their placement exams and remedial reading courses. This finding should inspire administrators to evaluate the motivation for these changes and consider the impact of those changes.

In these forecasting models, it is apparent that community college students have different risk factors to consider than traditional university students. Traditional risk factors such as standardized math scores or high school GPA have less bearing on the success of community college students, while enrollment status and reading comprehension may be more indicative of their future success.

Table 5. Stepwise Selection of Terms

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	24.269	8.0898	14.36	0.000
ACT read	1	2.377	2.3772	4.22	0.042
HS GPA	1	16.127	16.1274	28.63	0.000
age	1	4.827	4.8274	8.57	0.004
Error	125	70.422	0.5634		
Total	128	94.691			

Model Summary

S	R-sq(adj)	R-sq(pred)
0.750581	23.85%	20.45%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-1.409	0.770	-1.83	0.070	
ACT read	0.0273	0.0133	2.05	0.042	1.09
HS GPA	0.654	0.122	5.35	0.000	1.18
Age	0.0666	0.0228	2.93	0.004	1.09

Regression Equation

College GPA = -1.409 + 0.0273 ACT read + 0.654 HS GPA + 0.0666 age

Conclusion

This case study provided a chance to examine community college STEM student outcomes. This research indicates that the Mahalanobis Taguchi System can be used to identify important variables for forecasting completion of a STEM degree. The variables with large, positive S/N ratios were also included in the logistic regression model. This supports the use of MTS for pattern recognition and forecasting.

Based on this research, it appears that community college students have a different set of risk factors that could be used to predict their success in a STEM degree. Prior student performance as indicated by high school GPA did not appear to predict if a student will finish a STEM degree. A majority of previously published models showed a significance in high school GPA and math preparedness scores (Snyder & Cudney, 2017). This data was limited to student demographic data; therefore, there could be other factors to investigate to clearly understand the unique factors impacting completion rates among community college students.

While these initial results are promising, further research should be conducted to address several limitations of this study. Community colleges do not have admission standards; therefore, many applicants do not have standardized exam scores or report high school performance. The raw data is missing many important variables for students causing the sample to shrink considerably for the model.

The findings point to some areas of concern from the community college perspective. This is a time when many community colleges are scaling down their remedial reading courses, but reading aptitude appears to be a significant risk factor. Further research should be done to determine the exact impact reading ability has on a STEM student's ability to complete a degree. Additionally, research needs to be done on a more specific student performance scale. Are there courses that predict completion of a STEM degree? Does the starting point on the math pathway predict successful completion?

Future studies will further examine factors to build a stronger model for community college students. These risk factors are critical to community college student services. The only way to develop early alert systems is to have a more effective prediction model.

References

Adelman, C. (2005). Moving into Town--and Moving On: The Community College in the Lives of Traditional-Age Students. *US Department of Education*.

Carnevale, A. P., & Strohl, J. (2010). How increasing college access is increasing inequality, and what to do about it. In R. D. Kahlenberg (Ed.), *Rewarding Strivers: Helping Low-Income Students Succeed in College* (pp. 71-183). New York: The Century Foundation.

Carver, S., Van Sickle, J., Holcomb, J., Quinn, C., Jackson, D., Resnick, A., . . . Marquard, A. (2017). Operation STEM: increasing success and improving retention among first-generation and underrepresented minority students in STEM. *Journal of STEM Education, 18*(3), 20-29.

Chen, X. (2013). *STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001*. National Center for Education Statistics: U.S. Department of Education, Washington, D.C.

Cohen, Kisker, C. B., & Brawer, F. B. (2014). *The American community college* (Vol. Sixth;6;6th;). San Francisco, CA: Jossey-Bass, a Wiley brand.

Committee on Science, E., and Public Policy. (2007). *Rising Above The Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. (9780309100397;0309100399;). National Academies Press.

Costello, C. B. (2012). Increasing Opportunities for Low-Income Women and Student Parents in Science, Technology, Engineering, and Math at Community Colleges. Report# C388. *Institute for Women's Policy Research*.

Crisp, G., & Delgado, C. (2014). The Impact of Developmental Education on Community College Persistence and Vertical Transfer. *Community College Review, 42*(2), 99-117. doi:10.1177/0091552113516488

Ghasemi, E., Aaghaie, A., & Cudney, E. A. (2015). Mahalanobis Taguchi system: a review. *International Journal of Quality & Reliability Management, 32*(3), 291-307.

Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A.-B., & Handelsman, J. (2013). Increasing Persistence of College Students in STEM. *Science, 341*(6153), 1455. doi:10.1126/science.1240487

Hoffman, E., Starobin, S., Laanan, F. S., & Rivera, M. (2010). Role of Community Colleges in STEM Education: Thoughts on Implicaitons for Policy, Practice, and Future Research. *Journal of Women and Minorities in Science and Engineering, 16*(1), 85-96.

Horn, L., & Nevill, S. (2006). Profile of undergraduate students in US postsecondary education institutions: 2003–2004: With a special analysis of community college students (NCES 2006–184). Washington, DC: US Department of Education. *National Center for Education Statistics*.

Hurtado, S., Eagan, K., & Chang, M. (2010). Degrees of success: Bachelor's degree completion rates among initial STEM majors. *Higher Education Research Institute at UCLA, January*.

Olson, S., & Riordan, D. G. (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. *Executive Office of the President*.

Snyder, J., & Cudney, E. A. (2017). Retention Models for STEM Majors and Alignment to Community Colleges: A Review of the Literature. *Journal of STEM Education: Innovations and Research, 18*(3), 48-57.

Starobin, & Laanan, F. S. (2010). From Community College to PhD: Educational Pathways in Science, Technology, Engineering, and Mathematics. *Journal of Women and Minorities in Science and Engineering, 16*(1), 67-84. doi:10.1615/JWomenMinorScienEng.v16.i1.50

Terenzini, P. T., Lattuca, L. R., Ro, H. K., & Knight, D. B. (2014). America's Overlooked Engineers.

Woodall, W. H., Koudelik, R., Tsui, K.-L., Kim, S. B., Stoumbos, Z. G., & Carvounis, C. P. (2003). A review and analysis of the Mahalanobis—Taguchi system. *Technometrics, 45*(1), 1-15.