ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022
Paper ID #37780

Investigating Engineering Persistence through Expectancy Value Theory and Machine Learning Techniques

Campbell R Bego (Assistant Professor)

Dr. Campbell Bego researches engineering learning and persistence as an Assistant Professor in the Department of Engineering Fundamentals at the University of Louisville's Speed School of Engineering. Prior to entering academia, she obtained a BS in Mechanical Engineering from Columbia University, worked in tunnel ventilation (CFD modeling) at Mott MacDonald and AECOM, and received a Professional Engineering license in the State of New York. She draws on these experiences as well as her MS and PhD in Cognitive Science from the University of Louisville to construct meaningful activities in her first-year engineering course. She aims to improve the number of engineering graduates as well as the quality and diversity of the engineering workforce using evidence-based practices and applied theory in the classroom.

Pamela Bilo Thomas

Assistant Professor at the University of Louisville

Xiaomei Wang (Assistant Professor)

Investigating Engineering Persistence through Expectancy Value Theory and Machine Learning Techniques

Abstract

A vast amount of research has revealed that students decide to leave engineering for numerous reasons, including bias and discrimination, low performance in mathematics, belonging uncertainty, financial load, and individual interest. One possible framework to investigate students' persistence is Situated Expectancy Value Theory (SEVT), which proposes that students' achievement-related choices and performance are influenced by expectation of success, and subjective task value. In this contribution, we combined SEVT with data science tools to better understand students' academic decision-making, specifically the decision to persist in engineering. Our goal was to identify underlying patterns that predict persistence as well as illuminate possible interventions. Here, three machine learning tools, namely, clustering, principal component analysis (PCA), and decision trees, were applied to data from two cohorts of engineering students at a large public university. Concerning the SEVT framework, student responses to surveys given at the beginning and end of the first semester, containing established scales for self-efficacy and contingencies of academic competence self-worth (expectancies), and interest in engineering and perceived costs of studying engineering (subjective task values) were used. Demographic data including race, gender, and Pell eligibility, alongside performance data in the form of introductory course grades, GPA, and persistence into Year 2, complete the set of gathered information available to our data analyses. Collectively, we were interested in learning patterns that allowed us to use academic performance and SEVT data to predict engineering retention. Supporting previous findings in the literature, results from clustering analyses and a PCA indicate that performance in the first-semester engineering courses is highly predictive of persistence in engineering. Potential interventions include mid-first-semester feedback and learning interventions, as well as study habits and time management. New findings indicate that the SEVT framework also significantly predicts persistence, especially for mid-level performers. Discussions in the second semester with students are proposed for both research purposes as well as an intervention to improve student persistence.

1 Introduction

Technological innovation depends on a qualified and diverse engineering workforce [1, 2]. To remain internationally competitive, the US needs to improve recruitment, retention, and preparation of undergraduate engineering students, focusing particularly on improving the representation of underrepresented minorities [3]. This paper considers a broad range of factors that have been found to predict students' persistence through the first year of undergraduate engineering school,

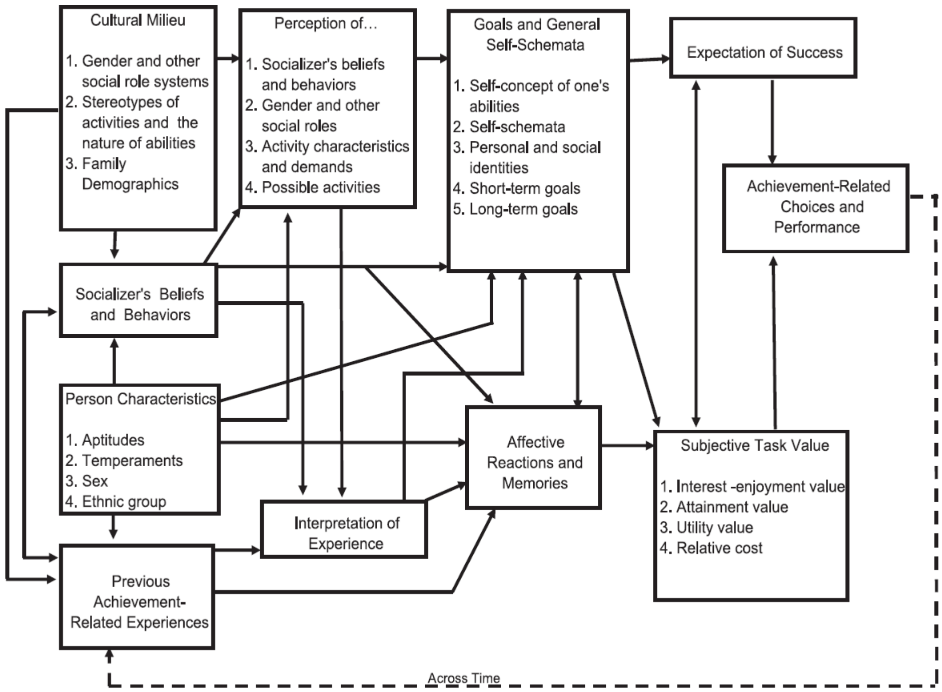

Figure 1: SEVT, as presented in [5]

with the goal of identifying potential interventions for improvement. The work is grounded in Situated Expectancy Value Theory (SEVT), which describes and relates the motivational factors behind students' academic decision-making [4, 5, 6]. Using data from multiple cohorts at a large public university, we apply machine-learning techniques to investigate the importance of demographic, performance, and noncognitive factors in the first year on student retention and grades.

1.1 Engineering Persistence and Situated Expectancy Value Theory (SEVT)

Research has revealed that students leave engineering for many reasons, including bias, discrimination, low performance in mathematics, belonging uncertainty, financial load, and individual interest [7, 8]. The number of significant predictors indicates that a holistic view of students' motivation is necessary to design the most effective interventions [9]. Some researchers have proposed persistence models specific to the student experience in engineering school, including Tinto (1987, [10]), Bean (2001, [11]), and Veenstra (2009, [12]). However, other more broadly applicable social cognitive theories and theories of motivation have been developed in recent years.

One framework is for describing student choices and performance is SEVT [4, 5, 6]. As visible in Figure 1, student behavior is influenced by two groups of factors: (a) expectation of success, and (b) subjective task value. According to leading SEVT researchers Wigfield and Eccles (2000, [6]), expectancies and values are based on social, cognitive, and task-specific beliefs such as belief in their own abilities, the difficulty of the task, and students' goals and identities. These variables are influenced by students' prior experiences and a variety of external social influences.

Each of these variables is broken down into several potential input variables, some of which are dynamic (e.g., "a student's perceptions of the possible activities") and some of which are static

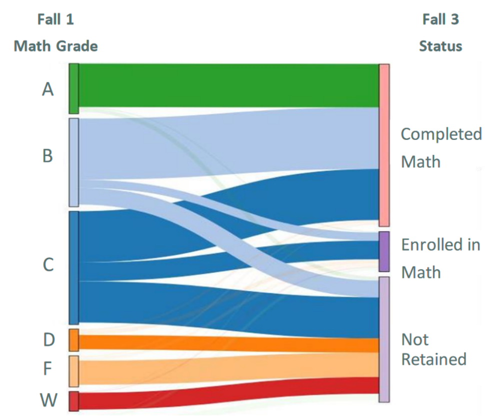

Figure 2: A ribbon plot of students' first-semester math grade and status in the third year. Most high-performing students completed the math sequence, which is a strong indication of future graduation. Most of the students who received D, F, or W grades in math were no longer in engineering by the third year.

(e.g.,"cultural milieu" and "personal characteristics"). The breadth and depth of each component varies widely, and each is meant to be interpreted and used for a given context (i.e., "situated" [5]). The connections between the components are also complex and dynamic.

SEVT has been useful in predicting both small-scale decisions, such as whether or not to engage in an in-class activity, and broad-scale decisions, such as persistence in an academic major. Perez et al. (2019 [13]), for example, identified three different latent SEVT profiles and examined achievement and course completion in STEM at the end of the first year of college. They found first that students with below-average competence beliefs and task values and above-average perceived costs were the least likely to be retained in STEM. They also found that underrepresented minorities were most commonly in a "moderate-all" profile, which was also associated with a lower average STEM GPA and fewer STEM courses completed.

1.2 The Current Study

Our goal is to use the SEVT framework to investigate students' academic decision-making, specifically the decision to persist into the second year at the engineering school within our large public university. The University of Louisville recruits a student body in which 25.2% are racial minorities, 33.7% are first-generation college students, and 37.7% are Pell-eligible (i.e., low income) [14]. Like other engineering schools with similar enrollment standards, we graduate fewer than 60% of our students within 6 years, and see even lower graduation rates for underrepresented minority students. Previous investigations have identified that most of the attrition occurs in the first year [15]. We have found that first-semester math grade predicts completion of the math sequence, which is directly related to a students' successful graduation in engineering (Figure 2).

An interdisciplinary group at the University of Louisville recently assembled a longitudinal database with performance, demographic, and survey responses. In 2018 and 2019, survey items included measurements for (1) self-efficacy [16] and (2) contingencies of academic competence (academic competence subscale [17]), (3) interest in engineering [18], and (4) perceived costs of engineering

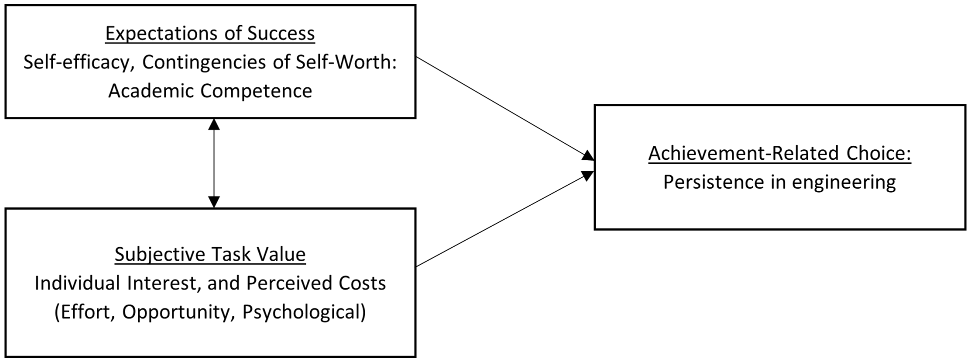

Figure 3: Primary decision-making factors for persistence within the SEVT framework.

school [19]. Scales 1 and 2 are expectancies, whereas scales 3 and 4 are subjective task values with respect to pursuing a degree in engineering. This is summarized in Figure 3.

We used clustering, principal component analysis (PCA), and decision trees to help us solve this problem. These algorithms helped us to learn about our students, and reduced the dimensionality of our dataset to identify the importance of the many available variables. The full list of performance, demographic, and survey variables are detailed below in the methodology.

Our research questions are as follows:

(1) Across academic performance in three key courses in the first year, ACT scores, and SEVT variables, which are the most important in predicting student persistence?

(2) How do variables from the SEVT framework influence student retention?

Lastly, we reflected on our results to develop potentially effective interventions.

2 Methods

This study was approved by the university's institutional review board.

2.1 Data

We retrieved de-identified data gathered from our engineering students over the 2018-2019 school years, restricting the dataset to those who were first-time full-time enrollment in fall of 2018 or 2019 (N = 995). Variables included:

1. ACT scores (composite, math, English and science/reading).
2. Responses to SEVT surveys, conducted at the beginning and end of the first semester, including: interest in engineering, perceived costs of studying engineering, self-efficacy, and contingencies of academic competence, academic competence subscale. Example items and references for each of these scales are provided in Table 1.
3. Academic performance in three required engineering courses that students usually take in their first term: math, chemistry. engineering fundamentals.

Table 1: SEVT Survey Items

Scale	Response Options	# Items	Example Item
Self-Efficacy [16]	1-Not at All True to 7-Very True	8	I'm certain I can understand the most difficult material presented in this course.
Contingencies of Academic Competence: Academic Competence [17]	1-Strongly Disagree to 7-Strongly Agree	5	Doing well in academics gives me a sense of self-respect.
Interest in Engineering [18]	1-Not at All True to 5-Very True	8	Engineering is practical for me to know.
Effort Cost [19]	1-Strongly Disagree to 6-Strongly Agree	4	When I think about the hard work needed to get through engineering school, I am not sure that it will be worth it in the end.
Opportunity Cost [19]	1-Strongly Disagree to 6-Strongly Agree	4	Studying for engineering school takes a lot of time away from other activities that I want to pursue.
Psychological Cost [19]	1-Strongly Disagree to 6-Strongly Agree	3	I'm concerned that my self-esteem will suffer if I am unsuccessful in engineering school.

4. Retention, defined as enrollment in the engineering school in the fall of the second year.

Collectively, we were interested in learning the relationship between 1-3 and 4 (or whether 1-3 predict 4). For details of variables and pre-processing, see Table 2.

2.2 Analysis

Three analysis methods were used for the current study: clustering analysis, principal component analysis (PCA), and decision tree classifier. For each analysis, records with missing data were excluded. Python packages sklearn[1] was used for programming.

Clustering analyses were conducted with the K-means algorithm [20, 21, 22] as the exploratory first step, using the elbow method [23] to decide the optimal number of clusters. Here we looked for sets of students that presented similar behaviors across sets of metrics (multiple course grades, ACT scores, beginning SEVT scores, and the change in SEVT values in the course of a student's freshman year). Each analysis identified clusters of students according to the selected variables,

[1] `https://scikit-learn.org/`

Table 2: Variables and Pre-Processing

1. ACT Score	
English, Math and Science	Standardized as 0-1
Composite Score	Mean of the separate scores, standardized as 0-1
2. SEVT Survey	
Self-efficacy	Likert Scale 1-5 standardized as 0-1
Academic Competence	Same as above
Interest in Engineering	Same as above
Perceived Cost of Engineering School	Same as above
3. First-term Academic Performance	
Math	Grade standardized to 4.0 scale F and W coded as 0
Chemistry	Same as above
Engineering Fundamentals	Same as above
4. Retention Result	
Retention	Dummy Coded, stayed as 0, left as 1

and we then compared these clusters using the persistence variable.

A PCA analysis was then conducted across all variables to try to reduce the dimensionality of the data by finding the input variables that best explain observed variability in the dataset [24]. The goal of a PCA analysis is to identify the principle component vectors that discriminate across all variables to predict student persistence. Because math performance relates so closely to persistence, we ran the the PCA method using the other input variables (students' chemistry grade, engineering fundamentals grade, ACT composite score, and beginning SEVT scores) to predict students' math grade. We attempted to see if we could cluster students into four categories: those that got an A or B in math, those that received a D, F, or W in math, those who got C's and persisted, or those who got C's and did not persist. We made the decision to divide the C students in that way, and not students who received the other grades, because C students presented an interesting analysis for us, as shown in Figure 2. It was our goal to see if we could find a signal that would help us predict which C students would leave, and which would stay.

Based upon the results from the clustering analyses and PCA analysis, we investigated our data further with decision tree analyses [25]. Decision tree analysis is an iterative procedure in which we seek cut-off values concerning the considered metrics (grades, SEVT and ACT scores) that would better explain or predict observed retention rates. We programmed our decision tree such that no leaf could contain less than 10 percent of our analyzed students. The dtreeviz[2] library was used for visualization of the decision tree, which was created using sklearn[3].

[2]`https://github.com/parrt/dtreeviz`
[3]`https://scikit-learn.org/stable/modules/tree.html`

3 Results

3.1 Demographic characteristics

Table 3 shows demographic characteristics of our sample.

Table 3: Demographic Characteristics

Category	N	%	Category	N	%
Sex			**Race**		
Male	778	78.19	White	797	80.10
Female	217	21.81	Asian	60	6.03
			Black/African American	46	4.62
			Hispanic/Latino	44	4.42
			Two or More Races	40	4.02
			Non-resident Alien	8	0.80
Pell-eligibility			**Birth Year**		
Not eligible	742	74.57	1999	196	19.70
Eligible	253	25.43	2000	494	49.65
			2001	290	29.15
			Other	15	1.51

3.2 Cluster analyses

Clustering analyses were performed separately across ACT scores, SEVT scores (at the beginning of the semester as well as the change over time), and first-semester engineering course performance. Results are presented in Figure 4, and described in the sections below.

3.2.1 Clustering across ACT scores

Table 4 shows student clusters based on ACT scores, including the number of students per cluster, the persistence rate of the cluster, and the average scores on the ACT components from the students on each of the derived clusters. As expected, the students with higher scores in all ACT components were most likely to persist (81% retained). Interestingly, the second highest rate of persistence was observed in cluster 3 (73% retained) which had high English and moderate math and science scores (i.e., English was their highest ACT score). The rate of persistence was higher than the cluster with moderate English but moderate-to-high math and science scores (i.e., math and science scores were higher than the English score). This is an interesting finding, and suggests the importance of English comprehension for first-year engineering students.

3.2.2 Clustering across SEVT scales

Table 5 presents results from surveys at the beginning of the fall semester, with scales normalized from 0 to 1. Table 5 shows six clusters of students, with significant variations in persistence (from 62% to 76%). Results indicate that students who initially presented relatively high scores in "Academic Competence", "Interest", and "Self-Efficacy" most likely stayed in the program despite also initially potentially perceiving a relatively high cost (see Table 5, cluster 6).

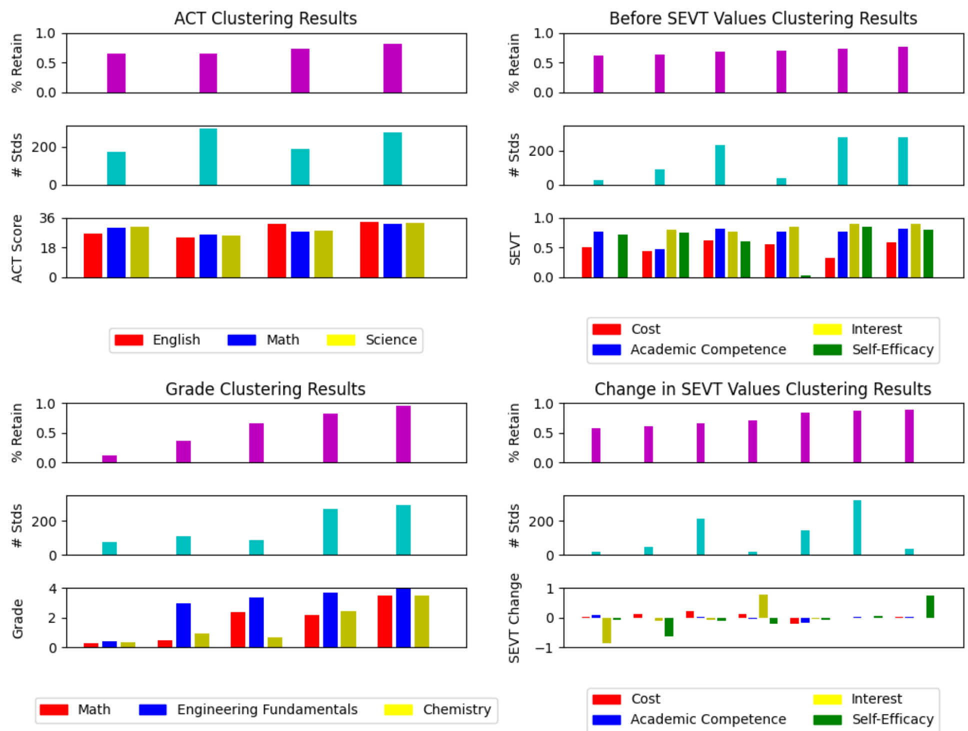

Figure 4: Student persistence and number of students for each of 4 clustering analyses.

Table 4: Clustering Analysis with ACT Scores

Cluster	# Students	% Persisted	English	Math	Science
1	175	64.57%	26.69	29.74	30.57
2	299	64.88%	23.99	26.13	25.06
3	189	73.02%	32.65	27.73	28.10
4	278	81.29%	33.72	32.49	33.24

Table 5: Clustering Analysis with Beginning SEVT Scores

Cluster	# Students	% Persisted	Perceived Costs	Academic Competence	Interest	Self-efficacy
1	29	62.06%	0.50	0.77	0.00	0.71
2	92	64.13%	0.43	0.46	0.80	0.74
3	232	68.53%	0.61	0.81	0.76	0.59
4	40	70.00%	0.55	0.76	0.84	0.02
5	279	73.47%	0.33	0.76	0.90	0.84
6	279	76.70%	0.59	0.82	0.89	0.79

Table 6: Clustering Analysis with Change in SEVT Scores

Cluster	# Students	% Persisted	Perceived Costs	Academic Competence	Interest	Self-efficacy
1	21	57.14%	0.01	0.07	-0.85	-0.05
2	49	61.22%	0.12	0.00	-0.09	-0.64
3	216	65.74%	0.21	0.02	-0.07	-0.11
4	21	71.42%	0.12	-0.04	-0.77	-0.19
5	144	84.02%	-0.19	-0.15	-0.04	-0.07
6	320	86.56%	-0.00	0.03	0.00	0.05
7	34	88.23%	0.03	0.03	-0.00	0.73

Table 6 presents results from difference in SEVT scores from the beginning to the end of the first semester, emulating the changes on the students perceived expectancies, values and costs due to their experiences throughout the academic semester. Here, scores can range from -1 to 1, where negative scores represents a decrease over time, and positive scores indicate gains. Table 6 shows seven clusters that also vary widely in rate of persistence (57% to 88%). The two clusters with the highest rates of persistence had either very little change across all variables, or a positive change in self-efficacy. The lowest rate of persistence occurred in a group with a large loss of interest in engineering (-0.85). Other changes associated with lower persistence rates were losses in self-efficacy (cluster 2), and gains in perceived costs (cluster 3).

3.2.3 Clustering across Engineering Course Grades

Table 7 identifies five clusters across student performance in their first-semester engineering courses. Students with the best performance were the most likely to persist (cluster 5, 96% persistence). The group of students with low performance in every course were the least likely to persist (cluster 1, 12%). Interestingly, clusters 3 and 4 illustrate that passing chemistry improves the chance that a C+ math student (indicated by the 2.37 and 2.20 grades, respectively) is retained.

Of the three courses required in the first semester, performance in the engineering fundamentals class seems to be the least variable. At our university, engineering fundamentals is designed to engage students in introductory engineering topics to help them understand the purpose and fundamental skills of engineering before delving into more complicated courses. It is therefore possible and expected for most students to do well. Math and chemistry grades seem to have more variance, indicating that they may be more difficult. It is interesting to note the difference between clusters 2, 3, and 4 alongside their rates of persistence. Cluster 2 has only 36% persistence, and has students with low performance in both math and chemistry. Persistence jumps up to 66% with passing performance in math, and then again to 82% with passing performance in both math and chemistry. This indicates that these may be considered ”barrier courses” at this university.

Table 7: Clustering Analysis with First-term Academic Performance

Cluster	# Students	% Persisted	Math	Chemistry	Engineering Fundamentals
1	77	11.69%	0.31	0.34	0.40
2	112	35.71%	0.44	0.93	2.96
3	89	66.29%	2.37	0.70	3.34
4	270	81.85%	2.20	2.46	3.72
5	292	95.54%	3.49	3.52	3.98

Due to the wide range of persistence rates in Table 7, we manually grouped the students by their grade in each course, and looked at the corresponding persistence rates. Results are presented in Table 8. As evidenced in the literature and prior work, these results indicate a strong correlation between first-semester grades and persistence in engineering.

Table 8: Percentage of student who stayed in engineering based upon first-term grades

Math	Number Stayed	Total	% Persisted
A	163	172	94.77%
B	243	282	86.17%
C	223	287	77.70%
DFW	75	254	29.53%
Chemistry	Number Stayed	Total	% Persisted
A	161	172	93.60%
B	218	247	88.25%
C	138	187	73.79%
DFW	97	251	38.64%
Engineering Fundamentals	Number Stayed	Total	% Persisted
A	528	600	88.00%
B	143	235	60.85%
C	18	52	34.61%
DFW	15	97	15.46%

3.2.4 Clustering Summary

Clustering results indicate that different groups of students with various combinations of incoming qualifications, perceptions, and performance in the first semester have different rates of persistence. One finding consistent with the literature is that high-performing students are more likely to persist in engineering. Although clustering across ACT and SEVT scores was not as effective in differentiating across persistence rates, these analyses have indicated that there may be some opportunities for interventions early in the first semester perhaps to improve performance, or at least student perceptions. Of importance, students with higher English ACT scores were more likely to persist, as were students who had little change in perceived expectancies or values across the semester or an increase in self-efficacy.

In summary, student clustering across grade data was best at predicting students' persistence. Because we had known from previous work that math performance was important for predicting student retention, as shown in Figure 2, we began to focus on math performance as both a predictor and an outcome of the noncognitive variables.

3.3 PCA analysis

We ran a PCA analysis using (1) beginning SEVT values, (2) ACT composite score, and (3) grades in engineering fundamentals and chemistry to see if we could identify principle components that predicted categories of math grades. Using the first two components, we drew Figure 5. Students with different math grades and retention outcomes were color coded. Vector results are shown in Table 9, and the explained variance is listed in Table 10.

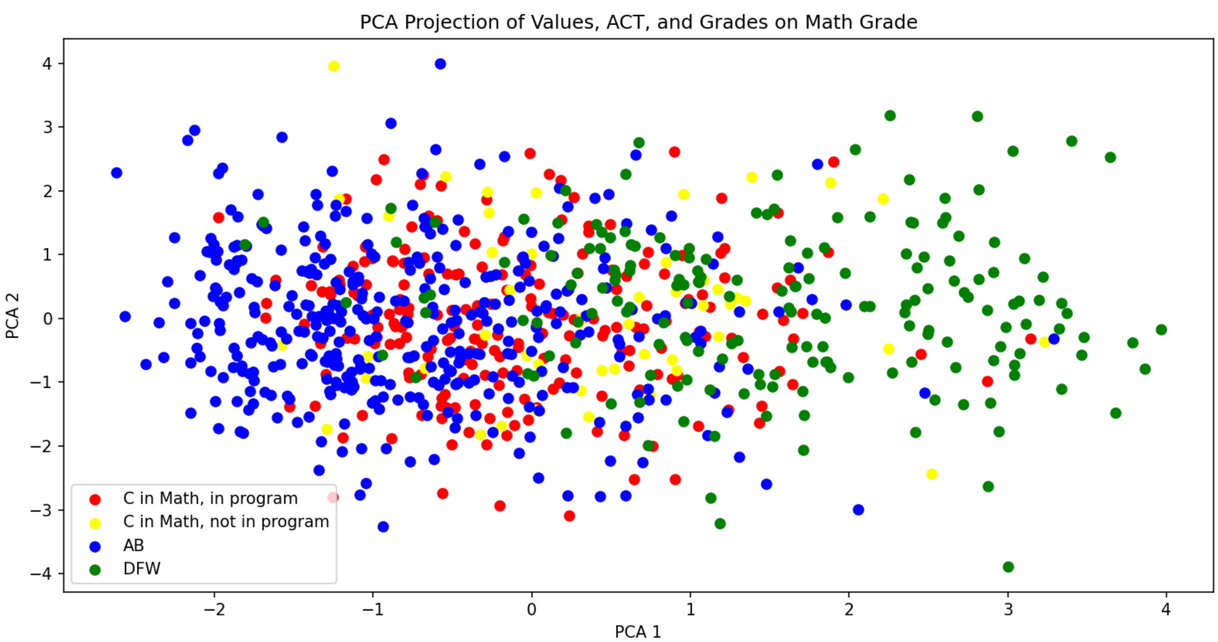

Figure 5: PCA results. Input variables were (1) beginning SEVT values, (2) ACT composite scores, and (3) grades in engineering fundamentals and chemistry, and the output variable was a category of math performance, as follows: (1) A or B in math, (2) C in math and persisted, (3) C in math and did not persist, (4) DFW in math.

The clear groupings visible in Figure 5 indicate that the two principle components did a decent job of differentiating between AB students and DFW students. However, students who received a C were not well differentiated. As shown in Table 10, the first two principal components were limited in their ability to explain all the variance in the dataset.

Table 9 presents interesting results about the rankings of the variables when determining the principal components. We see that much of the variance is explained in the first principal component by grades in chemistry and engineering (at 0.64 and 0.59) and ACT composite score (at 0.43). The second component consists instead of SEVT components, with perceived costs (0.64) and self-efficacy (0.55) as the most important variables. Then, in the third principal component, academic

Table 9: PCA Analysis Variable Loading

Variables	Component 1	Component 2	Component 3
First-term Chemistry	0.64	0.08	0.12
First-term Engineering Fundamentals	0.59	0.08	0.17
ACT Composite Score	0.43	0.09	0.09
Academic Competence	0.19	0.37	0.63
Self-efficacy	0.12	0.55	0.29
Cost	0.06	0.64	0.20
Interest	0.04	0.37	0.64

Table 10: Explained Variance

Component 1	Component 2	Component 3
0.26	0.20	0.14

competence (0.63) and interest (0.64) are the most important components. This suggests that in addition to the influence of academic performance, expectancies and subjective values of the students play an important role.

3.4 Decision Tree Analysis

In this analysis, we evaluated persistence using decision trees. As we mentioned earlier, we programmed our tree such that no leaf would contain less than 10 percent of the students we analyzed. We first used all data (grades, SEVT scores, and ACT scores) in the analysis to predict whether or not students persisted. Interestingly, the resulting decision tree (Figure 6, left) was one-sided, indicating that the splits at each level were strongly associated with retention without need for further input variables. The first split was on math grade, at 1.5 on the standard GPA scale, which is between a D+ and a C-. The second split was on engineering fundamentals, at a 3.66, which is approximately an A-. The third split was the chemistry grade of 2.5, which is between a C+ and a B-. Following these three primary splits, change in interest over time, ACT English score, and beginning level of self-efficacy. It is clear from these results that initial grades are the most important predictors of persistence.

Previous analyses identify the importance of math performance, as seen in Figure 6, but here, we looped back to our original question: How do variables from the SEVT framework influence students' decisions to remain in engineering school? Because what we saw from Figure 6, left, and Figure 2, we already concluded that the math grade is very important for identifying persistence. Therefore, we focused on students who received a C in math to identify how their SEVT values influenced their decision to stay in engineering. This is important because if we can identify which elements of the SEVT framework are important for our students, we can begin to design interventions to keep these students in school. The results are presented in Figure 6, right. It appears that change in interest in engineering appears to be the most important factor for students who decide to leave. Additionally, we see that the perceived cost of engineering school is another important factor for students to consider, because we see that split on multiple levels of our tree.

Decision Tree Analysis with Retention as an Outcome Variable

Figure 6: Left: Results from the decision tree analysis of all input variables (grades, ACT composite score, and SEVT scores), applied to the whole sample to predict persistence. Right: Analysis restricted to the students who received a C in math, with all input variables except for math grade. Green students persist, while yellow ones leave engineering.

3.5 Limitations

The university started using a standardized survey to collect student SEVT information from 2018 to present.

The school of engineering regulations for funding limited the persistence of students with low academic performance, potentially skewing some of our results.

4 Summary and Future Work

Supporting previous findings in the literature, results from clustering analyses and a PCA indicate that performance in the first-semester mathematics course is highly predictive of persistence in engineering (Table 7 & Table 8). In addition, passing performance is important in introductory chemistry, a first-semester required course, as well as a high grade in engineering fundamentals (A or B). These findings indicate that although calculus is often referred to as a barrier course, focusing on math performance only may limit the scope of interventions. In fact, one opportunity for intervention may occur in the middle of the first semester, specifically for students who appear to be falling behind in chemistry while maintaining average performance (C-level) in math.

New findings from our analyses indicate that the SEVT framework also significantly predicts persistence (Table 5). Although SEVT appears to be of secondary importance behind the first-semester grades (Table 9), the SEVT framework can also predict performance in the first math course (Table 9). It is therefore may be possible to design a predictive model using ML techniques and early surveys to identify a group of students on whom to intervene.

Lastly, decision tree analyses indicated that for the whole student population, the primary indicator of persistence following first-semester grades was performance in mathematics. The second most important was performance in chemistry, followed by performance in engineering fundamentals. The hierarchy of this tree shows the importance of first-semester grades, and how intervening inside or outside of the classroom to improve performance may improve persistence.

Once again, students who get a C in math are an interesting sample set, because these students are capable of becoming engineers, but may feel discouraged by their first-semester performance, which might not be due to their abilities but their transition to college life. For these students, decision tree analyses indicated that the change in interest over the first semester is the most important predictor of persistence. It is possible that students' interest dropped because they learned more about engineering and were no longer interested, but it is also possible that the lower interest was a result due to their underachievement. Other parts of the SEVT framework also help to identify likelihood of persistence. One interesting factor to consider in future research is the perceived costs of engineering school. It is possible that a different intervention timeline be used for students who receive a C in math, that could occur in the second semester. One on one discussions about achievement and effort with either peers or advisors could be both a research project as well as an effective retention intervention. Additionally, these students may be worried about financial aid eligibility, as is being researched concurrently with this paper. Either way, students who earn a C in math are an interesting population for targeted interventions. Insights from the early semester SEVT survey could likely be use to identify such students and properly design better approaches to reach this population.

References

[1] N. S. Board, "Revisiting the stem workforce: A companion to science and engineering indicators 2014," 2015.

[2] B. Hofstra, V. V. Kulkarni, S. M.-N. Galvez, B. He, D. Jurafsky, and D. A. McFarland, "The diversity–innovation paradox in science," *Proceedings of the National Academy of Sciences*, vol. 117, no. 17, pp. 9284–9291, 2020.

[3] National Science Foundation [NSF], "NSF, National Center for Science and Engineering Statistics. 2019. Women, minorities, and persons with disabilities in science and engineering," *Special Report NSF*, 2019.

[4] J. S. Eccles, T. Adler, R. Futterman, S. B. Goff, C. M. Kaczala, J. L. Meece, and C. Midgley, "Expectancies,

values, and academic behaviors," in *Achievement and achievement motivation*, J. T. Spence, Ed. San Francisco, CA: Free man, 1983, pp. 75–146.

[5] J. S. Eccles and A. Wigfield, "From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation," *Contemporary Educational Psychology*, vol. 61, p. 101859, 2020.

[6] A. Wigfield and J. S. Eccles, "Expectancy–value theory of achievement motivation," *Contemporary educational psychology*, vol. 25, no. 1, pp. 68–81, 2000.

[7] E. Seymour and N. M. Hewitt, *Talking about leaving*. Westview Press, Boulder, CO, 1997.

[8] M. Meyer and S. Marx, "Engineering dropouts: A qualitative examination of why undergraduates leave engineering," *Journal of Engineering Education*, vol. 103, no. 4, pp. 525–548, 2014.

[9] N. B. Honken and P. Ralston, "Freshman engineering retention: A holistic look," *Journal of STEM Education: Innovations and Research*, vol. 14, no. 2, 2013.

[10] V. Tinto, *Leaving college: Rethinking the causes and cures of student attrition.* ERIC, 1987.

[11] J. Bean and S. B. Eaton, "The psychology underlying successful retention practices," *Journal of College Student Retention: Research, Theory & Practice*, vol. 3, no. 1, pp. 73–89, 2001.

[12] C. P. Veenstra, E. L. Dey, and G. D. Herrin, "A model for freshman engineering retention." *Advances in Engineering Education*, vol. 1, no. 3, p. n3, 2009.

[13] T. Perez, S. V. Wormington, M. M. Barger, R. D. Schwartz-Bloom, Y.-k. Lee, and L. Linnenbrink-Garcia, "Science expectancy, value, and cost profiles and their proximal and distal relations to undergraduate science, technology, engineering, and math persistence," *Science education*, vol. 103, no. 2, pp. 264–286, 2019.

[14] "2021 just the facts," https://louisville.edu/facultyhandbook/Just&20the&20Facts.pdf, accessed: 2022-04-04.

[15] C. R. Bego, I. Y. Barrow, and P. A. Ralston, "Identifying bottlenecks in undergraduate engineering mathematics: Calculus i through differential equations," in *2017 ASEE Annual Conference & Exposition*, 2017.

[16] P. R. Pintrich *et al.*, "A manual for the use of the motivated strategies for learning questionnaire (mslq)." 1991.

[17] J. Crocker, R. K. Luhtanen, M. Lynne Cooper, and A. Bouvrette, "Contingencies of Self-Worth in College Students: Theory and Measurement," *Journal of Personality and Social Psychology*, vol. 85, pp. 894–908, 2003.

[18] Linnenbrink-Garcia *et al.*, "Measuring situational interest in academic domains," *Educational and psychological measurement*, vol. 70, no. 4, pp. 647–671, 2010.

[19] T. Perez, J. G. Cromley, and A. Kaplan, "The role of identity development, values, and costs in college stem retention." *Journal of educational psychology*, vol. 106, no. 1, p. 315, 2014.

[20] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[21] T. Kanungo *et al.*, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE transactions on pattern analysis and machine intelligence*, 2002.

[22] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, 2003.

[23] P. Bholowalia and A. Kumar, "Ebk-means: A clustering technique based on elbow method and k-means in wsn," *International Journal of Computer Applications*, vol. 105, no. 9, 2014.

[24] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2016.

[25] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142–147, 1977.