ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE

Paper ID #36794

Noncognitive Predictors of Engineering Persistence for C-in-Math Students: Exploring the Generalizability of Lasso Regression

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.

Jeffrey Lloyd Hieb

Jody Zhong (Graduate Assistant)

Jody is a doctoral student in the Counseling Psychology department at the University of Louisville College of Education & Human Development. Jody's interests include studying how students from underserved communities navigate academic and professional environments, and the formation of intersectional identities amongst undergraduate and graduate students.

Patricia A Ralston (Professor)

Thomas Tretter

© American Society for Engineering Education, 2022
Powered by www.slayte.com

Noncognitive Predictors of Engineering Persistence for C-in-Math Students: Exploring the Generalizability of Lasso Regression

Abstract

In this full-length paper, we present new research on engineering persistence for students who received a C in their first-semester math course. We implemented the least absolute shrinkage and selection operator ("lasso") method, a regularization technique, to consider the relative importance of several noncognitive variables known to impact engineering persistence. In addition to being a powerful tool for reducing dimensionality, lasso regression avoids model overfitting, a difficulty intrinsic to ordinary least squares linear regression, and therefore can reveal more generalizable results.

Our results indicated that students' perceived effort cost was the only predictive noncognitive factor in the fall 2019 cohort. Although this differed from our previous findings, both sets of data point towards situated expectancy-value theory as a powerful framework to explore engineering students' persistence. We also conclude that lasso regression is a useful tool in this domain, and additional studies are needed.

Keywords: *lasso regression, noncognitive predictors, persistence, interest, situated expectancy-value theory*

Engineering Persistence

For the last several decades, engineering educators have been striving to both understand and improve the rate of student persistence in engineering. As it stands, only approximately 55% of the students who enroll in engineering programs persist to graduation [1], [2]. Research has revealed that persistence is based on a wide variety of predictors from pre-college math preparation [3] to engineering school climate [4], engineering identity [5] and more. Several holistic models have been proposed over the years that have attempted to organize the factors [6]–[8], and these models have been instrumental in advancing our understanding of persistence and attrition for different student groups.

One key indicator of persistence appears to be math performance in the first semester of engineering school [9], [10]. While students who perform well typically continue towards successful graduation in engineering, students who do poorly in the first semester tend to leave. First-year math courses are prerequisite to discipline-specific engineering courses in following years. Therefore, students who fail early math courses could easily be discouraged from continuing in engineering, whereas very high performers (A's or B's) would have no trouble continuing on-track. The predictive models tell us that early performance is based on incoming traits and abilities as well as university climate and students' interactions with teachers and mentors, and that this performance impacts students' beliefs in their abilities and decisions to stay in engineering [6]–[8].

However, very little research has considered mid-level performers—those who receive a C in their first semester math course. At one large southeastern university, studies revealed that approximately equal numbers of students who receive a C grade in their first math course either persisted through their math requirements or left engineering by their third year (see Figure 1, [11]).

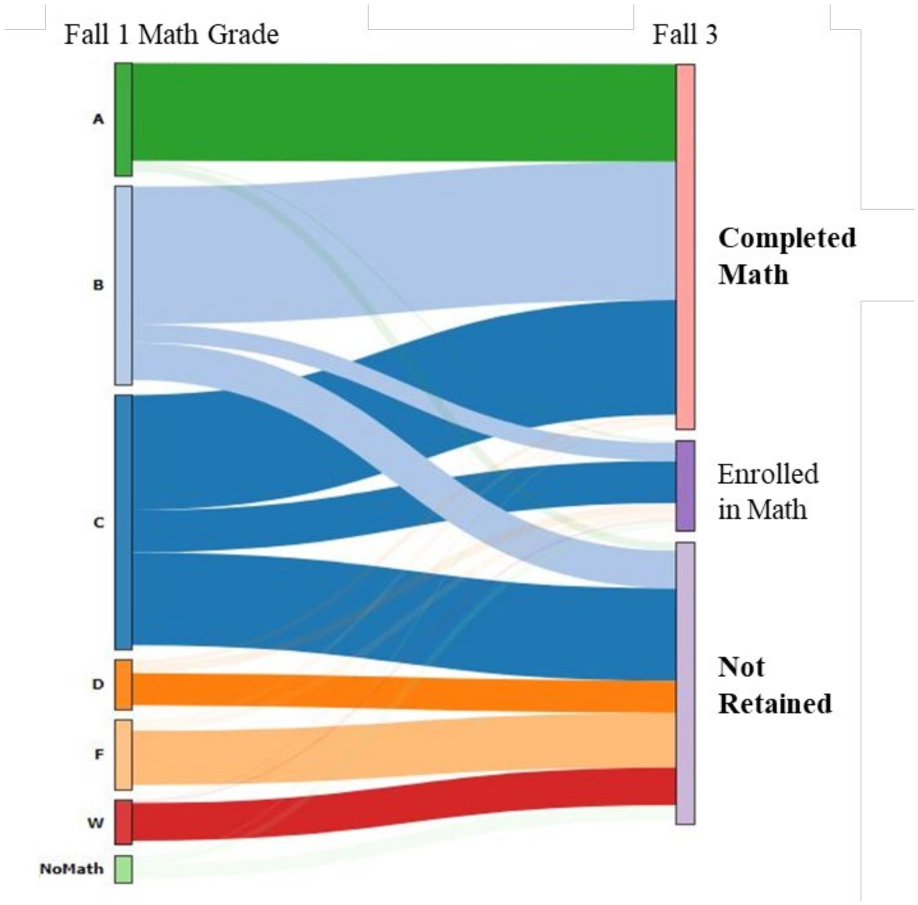

Figure 1: Initial math performance and corresponding math completion rates (a proxy for graduation rates) at the University of Louisville's J. B. Speed School of Engineering, from [11].

As reported in [11], the large number of C-in-math students who completed the math sequence indicates that it is possible to succeed in engineering with a mid-level grade and corresponding skills. Simultaneously, the group of C-in-math students who left engineering present an opportunity for interventions; there is a whole group of capable students who decide not to continue but are capable. Students with a C grade in the first semester math course therefore appear to be a potentially productive subgroup that merit further investigation for strengthening engineering retention.

In addition to math performance, several noncognitive variables such as a student's sense of belonging [12] or self-efficacy [13]–[15] have been found to predict student retention. These factors are interesting because not only can they be used to predict student decision-making, but they are frequently malleable and can change in response to interventions (see [16]). Again, many variables have been proposed and many have been found to significantly predict engineering student persistence. The difficulty is assimilating these possibilities to determine the most effective intervention strategy, particularly for an otherwise capable group of students.

Analysis Methods

Binary logistic regression is the most common quantitative method used to predict the outcome variable of engineering persistence from one or more independent variables (e.g., math performance). However, standard ordinary least squares (OLS) statistical methods generate a model that strongly fits the dataset at hand, and overfitting beta coefficients to a sample reduces generalizability of the results [17].

Regularization methods such as lasso regression tackle some of these limitations [18], [19]. In lasso regression, a penalty parameter (λ) is added to the model. The penalty parameter λ suppresses minimally predictive factors and lowers larger beta values, which reduces the explained variability but allows for results that are generalizable beyond the given dataset. The value of λ is what controls the amount of bias included in the model versus variance explained by the model. Therefore, the estimation of λ is important, and is determined iteratively, beginning with an extreme value of λ which suppresses all coefficients and moving towards a value of 0 (no penalty), which would result in a model equal to OLS regression.

For each value of λ, a model of best fit is estimated using cross-validation. In cross-validation, the data is split into *k*-folds, or *k* number of bins (*k* is typically 5 or 10). One fold is used for testing, and the remaining *k*-1 folds are used for training. A model is estimated for the training sample, and the predictive ability of the model is tested using the test sample. The process is then repeated, with each different fold serving as the test sample in turn. A measure of fit (e.g., mean squared error for linear models) is recorded for each fold *k* and averaged, for an overall measure of fit for a particular value of λ. Next, the same process is performed for a smaller value of λ. The iterative process looks for a value of λ that minimizes the averaged mean squared error, referred to as λ_{min}.

Grounding for the Current Study

Recently, our research group conducted lasso regression to investigate the relationship between a set of noncognitive factors (e.g., interest in engineering) and engineering persistence from the first to the second year for C-in-math students [20]. The analysis included students who matriculated in a single cohort at our large southeastern university (fall 2018). Only two of eight variables were identified by lasso regression as having coefficients greater than zero: (1) interest in engineering and (2) text anxiety.

In the current paper, we have run a lasso regression analysis on a second cohort (fall 2019) using the same noncognitive and outcome variables. Our research questions are as follows:

> *RQ1: What are the most important noncognitive predictors of engineering persistence for students who received a C in their first-semester math course in the fall 2019 cohort at the University of Louisville, J. B. Speed School of Engineering?*

RQ2: Are the most important variables for the fall 2019 cohort similar to those in the fall 2018 cohort? I.e., were our previous lasso analysis results generalizable?

The long-term goal of this research is to improve engineering student persistence. A replication in a second cohort will help us begin to understand the generalizability of our current and prior results. Our findings will also indicate where we could target future interventions for this academically at-risk population.

Methodology

Retrospective analysis was conducted on longitudinal data from our university. This analysis was approved by our Institutional Review Board.

Participants

Participants ($N = 113$) included all first-time full-time freshmen in fall 2019 at the University of Louisville's J. B. Speed School of Engineering who received a C in their first-semester mathematics course and responded to a survey during the first week of their freshman year.

Data

Longitudinal data included first-year retention status, first-semester math performance, and survey responses.

First-year retention status. For this analysis, a student was considered "retained" if they were enrolled at the engineering school in the summer 2020 semester. At the J. B. Speed School of Engineering, summer enrollment is expected, and nearly all students take summer classes. In the fall 2019 cohort, there were 92 C-in-math students who were retained and 21 C-in-math students who not retained.

Math performance. The engineering math courses at the J. B. Speed School of Engineering, are taught within the engineering school and include Calculus I, Calculus II, Calculus III, and Differential Equations. In addition, a preparatory math course (Introductory Calculus) is available for students who are underprepared for Calculus I. Engineering students therefore enroll in Calculus I or Introductory Calculus in their first semester at the engineering school. The current analysis included only students who received a C in one of these two courses.

Survey responses. Students responded to a first-year survey in the first week of fall 2019. Survey items were derived from established scales, measuring the noncognitive factors listed in Table 1.

Data Analysis Procedures

Lasso regression was performed in R, version 4.0.0. Principle libraries included `dplyr`, `psych`, and `glmnet`. First, data was pre-processed to reverse code all necessary items and compute scale averages. Descriptive and reliability statistics for all scales are summarized in Table 2. The random seed was set to 123 for consistency and repeatability. Lasso regression analysis was then performed using $k = 10$ fold cross-validation in the `glmnet` package. Lastly, significance of the coefficients was tested according to procedures in [21].

Table 1. Noncognitive scales – reference information.

Scale	Response Options	# Items	Example Item	Citation
Interest in Engineering	1-*Not at All True* to 5-*Very True*	8	Engineering is practical for me to know.	[22]
Effort Cost	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.	[23]
Opportunity Cost	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.	[23]
Psychological Cost	1-*Strongly Disagree* to 6-*Strongly Agree*	3	I'm concerned that my self-esteem will suffer if I am unsuccessful in engineering school.	[23]
Perceived Belonging Uncertainty	1-*Strongly Disagree* to 5-*Strongly Agree*	4	I am anxious about whether I fit in at college.	[16], [24]
Contingencies of Self-Worth: Academic Competence	1-*Strongly Disagree* to 7-*Strongly Agree*	5	Doing well in academics gives me a sense of self-respect.	[25]
Test Anxiety	1-*Not at All True* to 7-*Very True*	5	I have an uneasy, upset feeling when I take an exam.	(MSLQ) [26]
Self-Efficacy	1-*Not at All True* to 7-*Very True*	8	I'm certain I can understand the most difficult material presented in this course.	[26]

Table 2. Noncognitive scales - descriptive statistics and reliability.

Scale	M	SD	Cronbach's Alpha
Interest	4.33	0.52	0.84
Effort Cost	2.17	0.89	0.85
Opportunity Cost	3.54	1.24	0.88
Psychological Cost	3.89	1.34	0.86
Perceived Belonging Uncertainty	2.08	0.9	0.84
COSW: Academic Competence	5.34	0.97	0.77
Test Anxiety	4.19	1.34	0.85
Self-Efficacy	5.22	0.9	0.91

Results

The lasso regression procedures calculated an optimized penalty parameter of $\lambda_{min} = 0.055$. The binomial deviance for each λ value is plotted in Figure 2. The binomial deviance is the deviance in the cross-validated data, specifically: -2*the log-likelihood of the test data [27]. The minimum binomial deviance is when the penalty parameter, λ, results in the best predictive model based on cross-validation.

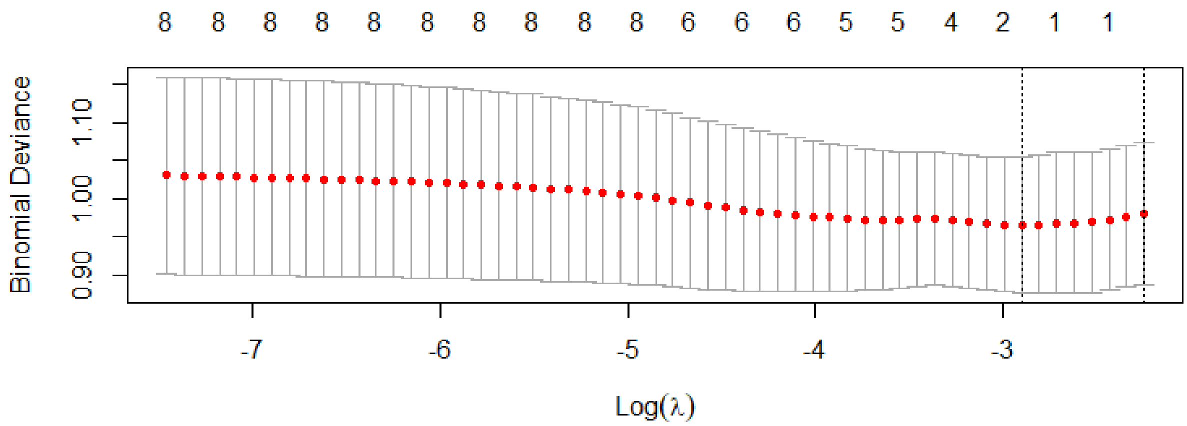

Figure 2. Binomial deviance for iterative λ values. The dashed line shows the location of λ_{min}.

The only significant coefficient in the model was for the Effort Cost factor, $\beta = -.33$, $p = .003$. The model's explanation of variance was rather low, with $R^2 = 0.05$, indicating that this model did not predict much of the variance in the data.

Discussion

Results indicate that in the fall 2019 cohort, the only noncognitive factor that was predictive of C-in-math student persistence was students' perception of the effort costs. Effort costs are defined as the perceived effort required to study engineering compared to other majors [23]. Costs are considered to be an important part of students' academic decision-making according to the situated expectancy-value theory of motivation (SEVT [28], [29]). The SEVT framework, very briefly, proposes that a student makes decisions based upon their expectations of success (e.g., their self-efficacy, as well as observation of their previous performance), and subjective task value (i.e., their interest in engineering, after considering the costs of studying engineering). The significant coefficient was negative, indicating that students who perceived higher effort costs were less likely to be retained. As it is a common perception that engineering school requires more effort than other majors, our findings are reasonable and consistent with recent literature in STEM fields [23].

However, our previous findings from an earlier cohort indicated that interest in engineering and test anxiety were the most important factors for C-in-math student persistence [20]. It is worth mentioning that interest is also a factor in SEVT, and alongside perceived costs, it adds to students' subjective task value. Our two sets of findings may in fact be pointing to the same decision-making factor, although the important subfactors were different for our two cohorts.

In retrospect, an obvious difference between fall 2018 and fall 2019 was the condition of the summer semester following students first year; fall 2019 students were forced to enroll in remote classes in their first summer semester (summer 2020) due to COVID-19. It is therefore very possible that the students who stopped engineering school from the fall 2019 cohort had different reasons for withdrawing than those in the fall 2018 cohort. With this difference in mind, it is an interesting finding that students' perceptions of effort costs were the most significant predictor of student persistence. Those who came into engineering school expecting it to require additional effort, when faced with the additional challenges and effort necessitated because of COVID-19 restrictions, were more likely to leave than those who perceived lower effort costs.

Although it is possible to conclude that our results did not replicate and therefore the lasso method is not generalizable, we propose to interpret our combined results as slowly uncovering the validity and usability of the SEVT framework for predicting student persistence. It may well be that the positive effect of a stronger interest in engineering (previous finding) could be somewhat protective against a potential negative effect of perceived high effort cost (current finding). But for our current study cohort, the additional negative effort and challenge imposed in summer 2020 by COVID pandemic restrictions may have simultaneously weakened existing interest and/or magnified perceived effort cost. We believe more work is needed, but it is likely that an intervention that attempts to increase students' valuation of obtaining an engineering degree may counteract perceived higher effort cost and subsequently improve persistence for students who earn a C in their first math course.

Limitations

Although we used a methodology that has been established in small datasets, it should be recognized that there were $N = 163$ students who were included in the prior study, and $N = 113$ students included in the current study, only 21 of whom did not persist beyond their first year. The results of this study are tied directly to those students, who may differ from others in different years at our university and who differ from the population at large. Future work will extend this into even more cohorts, and potentially other universities.

Conclusions

For the fall 2019 cohort at our southeastern university, the only significant noncognitive predictor of C-in-math student persistence was perceived effort costs. Although this finding did not replicate the importance of interest and test anxiety in our previous paper, both sets of results fit within situated expectancy-value theory (SEVT) and are consistent with the literature at large. We believe that lasso regression is an effective tool for analyzing engineering persistence due to the complex, multivariate models needed alongside a limited number of datapoints. Results from these lasso regression analyses may prove to have stronger generalizability compared to other regression approaches, which can be tested in future cohorts to explore if SEVT continues to serve as a useful framework for interpreting patterns of results over time.

References

[1] G. Zhang, T. J. Anderson, M. W. Ohland, and B. R. Thorndyke, “Identifying factors influencing engineering student graduation: A longitudinal and cross-institutional study,” *J. Eng. Educ.*, 2004, doi: 10.1002/j.2168-9830.2004.tb00820.x.

[2] M. W. Ohland, S. D. Sheppard, G. Lichtenstein, O. Eris, D. Chachra, and R. A. Layton, “Persistence, engagement, and migration in engineering programs,” *J. Eng. Educ.*, vol. 97, no. 3, pp. 259–278, 2008.

[3] W. Pearson Jr and J. D. Miller, “Pathways to an engineering career,” *Peabody J. Educ.*, vol. 87, no. 1, pp. 46–61, 2012, doi: 10.1080/0161956X.2012.642270.

[4] M. Crawford and M. MacLeod, “Gender in the college classroom: An assessment of the ‘chilly climate’ for women,” *Sex Roles*, vol. 23, no. 3–4, pp. 101–122, 1990, doi: 10.1007/BF00289859.

[5] A. D. Patrick, M. Borrego, and A. N. Prybutok, “Predicting persistence in engineering through an engineering identity scale,” *Int. J. Eng. Educ.*, vol. 34, no. 2a, 2018.

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

[7] C. P. Veenstra, E. L. Dey, and G. D. Herrin, “A model for freshman engineering retention,” *Adv. Eng. Educ.*, vol. 1, no. 3, 2009.

[8] J. Bean and S. B. Eaton, “The Psychology Underlying Successful Retention Practices,” *J. Coll. Student Retent. Res. Theory Pract.*, vol. 3, no. 1, pp. 73–89, 2001, doi: 10.2190/6r55-4b30-28xg-l8u0.

[9] J. Van Dyken, L. Benson, and P. Gerard, “Persistence in engineering: does initial mathematics course matter?,” in *2015 ASEE Annual Conference \& Exposition*, 2015, pp. 26–1225.

[10] J. L. M. Wilkins, B. D. Bowen, and S. B. Mullins, “First mathematics course in college and graduating in engineering: Dispelling the myth that beginning in higher-level mathematics courses is always a good thing,” *J. Eng. Educ.*, vol. 110, no. 3, pp. 616–635, 2021.

[11] C. R. Bego, J. L. Hieb, and P. A. S. Ralston, “Barriers and bottlenecks in engineering mathematics: How performance throughout a math sequence affects retention and persistence to graduation,” in *October 2019 IEEE Frontiers in Education Conference (FIE)*, 2019.

[12] J. A. Morrow and M. Ackermann, “Intention to persist and retention of first-year students: The importance of motivation and sense of belonging.,” *Coll. Stud. J.*, vol. 46, no. 3, pp. 483–491, 2012.

[13] A. Bandura, “Social cognitive theory: An agentic perspective,” *Asian Journal of Social Psychology*. 1999, doi: 10.1111/1467-839X.00024.

[14] F. Wu, W. Fan, C. Arbona, and D. de la Rosa-Pohl, “Self-efficacy and subjective task

values in relation to choice, effort, persistence, and continuation in engineering: an Expectancy-value theory perspective," *Eur. J. Eng. Educ. TA - TT -*, vol. 45, no. 1, pp. 151–163, 2020, doi: 10.1080/03043797.2019.1659231 LK - https://uofl.on.worldcat.org/oclc/8508703012.

[15] J. P. Concannon, S. B. Serota, M. R. Fitzpatrick, and P. L. Brown, "How Interests, self-efficacy, and self-regulation impacted six undergraduate pre-engineering students' persistence," *Eur. J. Eng. Educ. TA - TT -*, vol. 44, no. 4, pp. 484–503, 2019, doi: 10.1080/03043797.2017.1422695 LK - https://uofl.on.worldcat.org/oclc/8192207321.

[16] G. M. Walton and G. L. Cohen, "A brief social-belonging intervention improves academic and health outcomes of minority students," *Science (80-.).*, vol. 331, no. 6023, pp. 1447–1451, 2011, doi: 10.1126/science.1198364.

[17] D. M. Hawkins, "The problem of overfitting," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, 2004.

[18] D. M. McNeish, "Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences," *Multivariate Behav. Res.*, vol. 50, no. 5, pp. 471–484, 2015, doi: 10.1080/00273171.2015.1036965.

[19] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc. Ser. B*, no. 58, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[20] C. R. Bego, J. C. Immekus, and J. L. Hieb, "Predictors of first-year retention among undergraduate engineering students who earned a C in their first-semester mathematics course," in *Proceedings of the 127th ASEE Annual Conference and Exposition*, 2020, pp. 1–10.

[21] W. H. Finch and M. E. H. Finch, "Regularization methods for fitting linear models with small sample sizes: Fitting the Lasso estimator using R," *Pract. Assessment, Res. Eval.*, vol. 21, no. 1, p. 7, 2016.

[22] L. Linnenbrink-Garcia *et al.*, "Measuring situational interest in academic domains," *Educ. Psychol. Meas.*, vol. 70, no. 4, pp. 647–671, 2010, doi: 10.1177/0013164409355699.

[23] T. Perez, J. G. Cromley, and A. Kaplan, "The role of identity development, values, and costs in college STEM retention," *J. Educ. Psychol.*, vol. 331, pp. 1447–1451, 2014, doi: 10.1037/a0034027.

[24] J. P. Weaver, M. S. DeCaro, J. L. Hieb, and P. A. Ralston, "Social belonging and first-year engineering mathematics: A collaborative learning intervention," 2016, doi: 10.18260/p.25818.

[25] J. Crocker, R. K. Luhtanen, M. Lynne Cooper, and A. Bouvrette, "Contingencies of Self-Worth in College Students: Theory and Measurement," *J. Pers. Soc. Psychol.*, vol. 85, pp. 894–908, 2003, doi: 10.1037/0022-3514.85.5.894.

[26] P. R. R. Pintrich, D. Smith, T. Garcia, and W. McKeachie, "A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)," Ann Arbor, MI, 1991.

[27] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear

models via coordinate descent," *J. Stat. Softw.*, vol. 33, no. 1, p. 1, 2010.

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

[29] J. S. Eccles *et al.*, "Expectancies, values, and academic behaviors," in *Achievement and achievement motivation*, J. T. Spence, Ed. San Francisco, CA: Free man, 1983, pp. 75–146.