



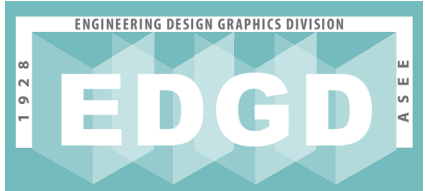
2025



Volume 88



THE ENGINEERING DESIGN GRAPHICS  
**Journal**





# Table of Contents



<b>Table of Contents</b> .....	i
<b>Editorial Board, Advisory Board, and Review Board</b> .....	ii
<b>EDGD Calendar of Events</b> .....	iii
<b>Message from the Editor</b> .....	iv
<b>Understanding Factors of Engineering Student Persistence Using Predictive Modeling</b> ....	1
Daniel P. Kelly, Erik J. Schettig, Aaron C. Clark and Jeremy V. Ernst	

 **Editorial Board ■ Advisory Board ■ Review Board** 

**EDGD Chair**

**Erik Schettig**, North Carolina State University

**Editorial Board**

**Editor:** Nancy E. Study, Penn State Behrend

**Layout Editor:** Judith A. Birchman, Purdue University

**Photographer:** Theodore Branoff, Illinois State University

**Circulation Manager:** Nancy E. Study, Penn State Behrend

**Advisory Board**

**Judith A. Birchman**, Purdue University

**Robert A. Chin**, East Carolina University

**La Verne Abe Harris**, Arizona State University

**Review Board**

**Jody L. Alberd**, Austin Peay State University

**Ron Barr**, The University of Texas at Austin

**Isaac Chang**, Illinois State University

**Aaron Clark**, North Carolina State University

**Cameron Denson**, North Carolina State University

**Nate Hartman**, Purdue University

**Tyler Kreipke**, University of Portland

**Jeff Morris**, Rensselaer Polytechnic Institute

**Steven Nozaki**, Penn State Behrend

**Clodagh Reid**, Athlone Institute of Technology

**Rahul Renu**, Francis Marion University

**Shaobo Huang**, University of Saskatchewan

**Heidi Steinhauer**, Embry-Riddle Aeronautical University

**Mostafa Tossi**, Penn State Scranton

**Online Distribution**

The online EDGJ is a reality as a result of support provided by **East Carolina University** and **Biwu Yang**, Research & Development, ECU Academic Outreach.

# EDGD Calendar of Events

## Future ASEE Engineering Design Graphics Division Midyear Conferences

**EDGD Midyear Conference** – To be discussed at the ASEE Annual Conference

## Future ASEE Annual Conferences

Year	Dates	Location	Program Chair
2026	June 21 - 24	Charlotte, North Carolina	
2027	June 20 - 23	Toronto, Ontario, Canada	

If you're interested in serving as the Division's program chair for any of the future ASEE annual conferences, please contact the Director of Programs, Erik Schettig

# Message from the Editor

**Nancy E. Study**, *EDGJ* Editor

Penn State Behrend

My last few letters from the editor have been focused on the future of the Journal and posing questions about what to do to improve the number and quality of submissions. We held meetings at the ASEE Annual Conferences, solicited feedback in a variety of other venues, and I am pleased that we are making small steps forward. That said, we still need your help in encouraging your colleagues and graduate students to submit articles that meet our expanded scope. Please refer to the information on our website and look for our calls for manuscripts on the EDGD listserv and other related listservs.

I had previously notified our readers that, as a result of the lack of regular publication, we were no longer indexed by ERIC. Because of changes in US government policy and in ERIC, we are not pursuing getting indexed again at this time, but may do so in the future if policies change, and we are able to maintain a regular publication schedule.

I would like to announce that under the umbrella of ASEE, we now have DOI numbers associated with Journal Volumes. This was via the efforts of our work on the ASEE Committee on Scholarly Publications, and the help of the staff at the Joyner Library at East Carolina University, who host our Journal online.

The Journal staff welcome your feedback as we continue to move forward in the changing environment of academia and publishing. As always, I thank Judy Birchman for her able assistance in doing the copy-editing, the staff at East Carolina University, and all the Journal's reviewers for their timely feedback. Hope to see you all in Charlotte.

## Understanding Factors of Engineering Student Persistence Using Predictive Modeling

Daniel P. Kelly, Erik J. Schettig and Aaron C. Clark

*North Carolina State University*

Jeremy V. Ernst

*Embry Riddle Aeronautical University*

### Abstract

Graduation rates of undergraduate engineering programs are low in comparison to other majors. Historically, engineering programs graduate between 50% and 55% of the students who initially matriculate. This average is even lower for women and underrepresented minorities. There is a large body of research that discusses the reasons students choose engineering and why they leave. Missing in this literature is an in-depth mathematical analysis of the factors related to persistence and non-persistence and how that knowledge can be used to improve programs, courses, and textbooks with the explicit goal of increasing persistence. This paper summarizes the techniques and findings from the use of machine learning software to create predictive models that are used to determine the most influential factors of student persistence in engineering programs from a study conducted in Fundamental Engineering Graphics courses. Dozens of variables including academic scores, non-cognitive and skill-based assessments, and demographic information for 300 students were used to develop a model capable of predicting whether a student will persist with nearly 94% accuracy. This research determined that age, gender, self-efficacy, and whether a student's parents were engineers were the most influential factors of persistence. Using this information, combined with the theoretical underpinnings of the constructs, this paper suggests changes to current practice to specifically target these factors to improve persistence rates in engineering education.

### Introduction

Some years ago, the deceleration in science and engineering advances were highlighted as focal concerns by the National Research Council. These concerns span those of future security and economic vibrance (National Research Council, 2007 as cited by Lockhart and Rambo-Hernandez, 2023). In a swift response, many colleges and universities within the United States began the prioritization of engineering degree programs and engineering workforce preparation just to quickly encounter student attrition concerns. According to the American Society of Mechanical Engineers (Flanigan, 2024), on a national scale only half of engineering majors graduate. This rate has largely been the norm for several decades (Bestfield-Sacre, Atman, Shuman, 1997).

Attrition can partially be attributed to academic readiness and performance (Amin, Sanders, Kidd,

& Rambo-Hernandez, 2024), however, beyond competency-based outcomes are factors serving as indicators of retention and success in engineering. For example, there is a greater likelihood of students persisting in engineering when they both perceive themselves as capable of succeeding, as well as have an appreciation and understanding of the field (Lee, et al, 2022). This creates a need to account for success and retention factors to build enhanced mechanisms to address non-persistence.

A more exacting mechanism to detect and identify students at risk of non-persistence is necessary in order to be proactive and to structure interventions appropriately. Efforts to accurately identify students at risk have been approached with many different models based on predictor variables (Mason, Twomey, Wright, and Whitman, 2018). Large-scale, multiple-variable analyses are needed to identify the composite factors



**Table 1**  
*Included Variables.*

<b>Variable</b>	<b>Collection</b>	<b>Citation</b>
Race/Ethnicity	Pre-Test	Kelly, et al., 2019
Gender	Pre-Test	Kelly, et al., 2019
GPA	Pre-Test	Kelly, et al., 2019
Age	Pre-Test	N/A
First Generation College Student	Pre-Test	Kelly, 2020
Parents are Engineers	Pre-Test	N/A
Mental Rotation	Pre- and Post-Test	Sorby & Baartmans, 2000
Self-Regulation	Pre- and Post-Test	Ernst & Clark, 2014
3D Modeling Self-Efficacy	Pre- and Post-Test	Denson et al., 2018

generally, and engineering specifically, are vast and difficult to quantify. Personal, social, academic, financial, health, background, and motivation can all influence the likelihood of a student persisting. Additionally, these factors are influenced in ways unique to different populations of students, creating heavily nested sub-groups. However, the nested groupings were not readily apparent given the number of variables assessed, the variations within those variables, and at what point in continuous variables were differences in persistence rates apparent, if not significant. Similarly, persistence rates based on categorical variables were not accurate due to nested groups and a lack of sensitivity for non-binary variables. For example, logistic regression found several significant factors such as the pre-test three-dimensional modeling self-efficacy (3DSE) score, gender, and whether the student's parents were engineers, but could not capture for which sub-groups of students the factors were most significant. Correlation analysis produced similar results with statistically significant associations—however, these were weak with a maximum coefficient of .147—between persistence and mental rotation (post-test), parents who are engineers, registration with disability services, GPA, and gender. Creating further confusion, many of the

variables are binary so which group the associations favor is not known.

Although it may be technically possible to create a model for every potential sub-group and examine the probability of persistence, this would be exceptionally inefficient and lead to a high probability of type I and II errors. Additionally, not knowing where classification splitting would occur for continuous variables (and that those values would likely be different for each sub-group) makes this method impossible. To overcome these barriers, data mining and visualization software were employed to examine the effects of the variables and their values on persistence and classify groups for which distinct variables had the greatest effect.

### **Classification**

To better explore which variables, and relationships between them, affected the persistence in engineering education among unknown sub-groups, data mining software was employed. The software chosen, Orange (Demšar et al., 2013), provided an open-source solution that was capable of running the analysis and appropriate for exploratory data analysis and visualization (Saragam, K. & Vivekanandam, 2018). Data mining



software, such as Orange™, is often used in predictive analysis and producing machine learning algorithms that examine several variables and classify objects based on the characteristic values entered. A classic example of this is the classification of flowers based on height, number of petals, length of leaves, etc. Based on these characteristics, a model is created that calculates the likelihood of a particular specimen being a particular flower. This is done repeatedly to train the algorithm to identify future observations without having the type of flower listed.

For this study, a classification and regression tree (CART) model was employed. A CART model does not require predefined relationships between the dependent and independent variables (Chang & Wang, 2006). CART does not require strong model assumptions and is absent the need for predetermined relationships between variables (Chang, Lin, Kwok, & Saw, 2023), making it a suitable method to examine the reasons students and sub-populations of students persisted in engineering education among our study population.

Due to the exploratory nature of this study, a classification tree was chosen to examine the

variables that either defined sub-groups or explained which variable had the greatest association with persistence, or non-persistence. Figure 1 displays the classification tree as displayed by the software. The colors indicate the strength of the relationship between the variable and persistence with red indicating persistence and blue indicating non-persistence. The shade of the color or the strength of the relationship with the nodes changing to white as the proportion approaches fifty percent. This example includes all the variables described previously.

At its core, a classification tree uses an algorithmic approach, with no deference to underlying knowledge or theory, to split the observations into binary categories based on the variable values in each observation. In this case, the binary classification is persistence in engineering [Yes] and non-persistence [No]. For categorical variables, all binary and dummy coded in this study, the split is simply which variable value exists within a particular observation. For continuous variables, the algorithm employs a regression analysis which determines the splitting point at its most mathematically logical point rather than that of a predetermined value. The advantage of this model is that it is adaptive and will self-ad-

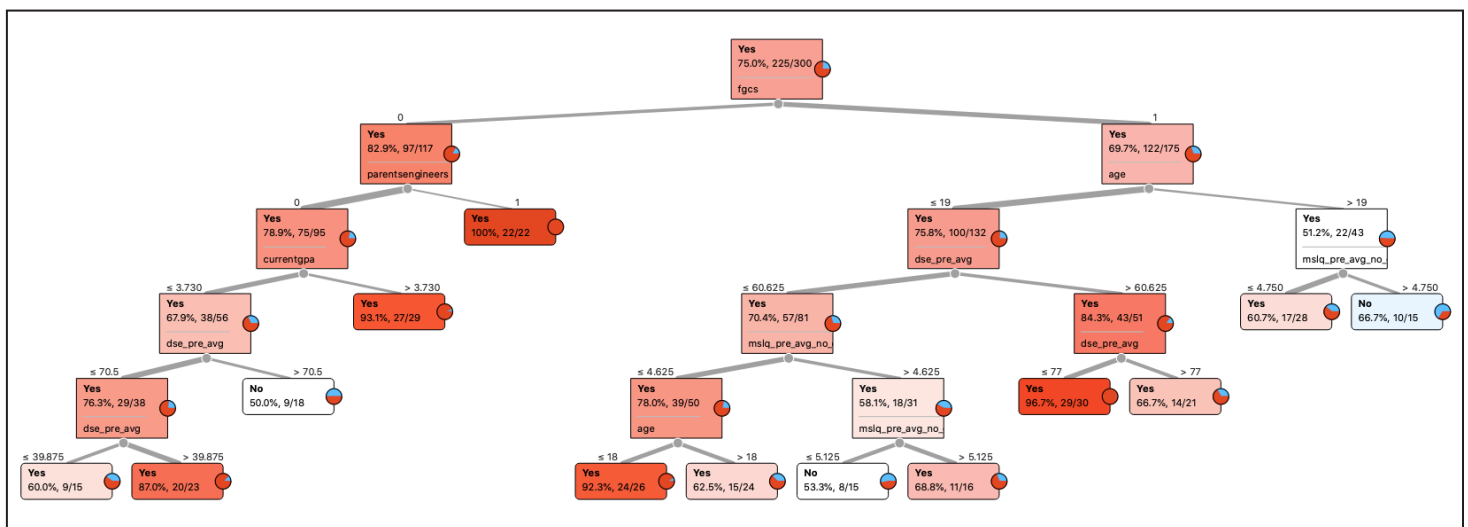


Figure 1. Example of a classification tree derived in orange.



FGCS status, gender, parent(s) is/are engineers, age, and 3DSE (pre-test). Most of the removed variables had no effect on the classification tree indicating that the remaining factors of persistence are of significant influence.

Table 4 displays the classification accuracy of the pruned classification tree with the revised values of the other models discussed previously. In addition to a higher classification accuracy, the classification tree is also significantly closer to the highest accuracy model (SVM), closing the original gap by 55% and moving from a rank of four to a rank of three.

**Table 4**  
*Original Classification Models Tested and Their Accuracy.*

Model	Accuracy
Support-Vector Machines (SVM)	.767
Logistic Regression	.742
<b>Classification Tree</b>	<b>.741</b>
Random Forest	.739
k-Nearest Neighbor (kNN)	.701

With the 3.8% overall increase in accuracy, the pruned model is able to accurately predict 89.5% of students who persist and 28% of students who do not. Although lower, there is still a significant percentage of false positives with 69.5% of misclassifications falling into this category. Table 5

**Table 7**  
*Classification Models Tested and Their Accuracy for Each Variable Set.*

Model	Accuracy (12 variables)	Accuracy (5 variables)	Accuracy (4 variables)
Classification Tree)	.703	.741	<b>.759</b>
Logistic Regression	.737	.742	.750
Support-Vector Machines (SVM)	<b>.750</b>	<b>.767</b>	.741
Random Forest	.716	.746	.728
k-Nearest Neighbor (kNN)	.694	.701	.702

displays the percentages of accuracy for the actual outcome for the pruned tree.

**Table 5**  
*Confusion Matrix for Pruned Classification Tree (Five Variables).*

		Predicted	
		No	Yes
Actual	No	<b>28.0%</b>	72.0%
	Yes	10.5%	<b>89.5%</b>

In the original model, displayed in Figure 1, all variables were included figure in the classification tree. Pruning reduced the number of variables to five. At the root node of both trees was FGCS status. For this reason, FGCS was retained as a variable. However, the removal of this variable increases the model's classification accuracy to 75.9% and increases the model's ability to predict student persistence to 93.8% but reduces the model's ability to predict non-persistence to 22.2% as shown in Table 6.

**Table 6**  
*Confusion Matrix for Pruned Classification Tree (Four Variables).*

		Predicted	
		No	Yes
Actual	No	<b>22.2%</b>	77.8%
	Yes	6.2%	<b>93.8%</b>

Removing the FGCS variable also makes the classification tree the most accurate predictive model of the five models compared. Table 7 displays

the classification accuracy of all five models for each of the three variable sets.

As it is the root node in both models FGCS status clearly represents an important variable in persistence; however, it is unclear why its removal increases model accuracy. There are reasons for both the inclusion and removal of the FGCS variable from the models. However, as inclusion increases overall model accuracy and persistence prediction, but removal increases Type I error, as well as the existence of similar trends in both pruned classification trees, FGCS status is included in the discussion.

## Discussion

Although a predictive model was used to identify the factors of persistence in engineering education, the prediction of persistence is not the goal of this study. Additionally, the remaining variance, false positives, false negatives, and the sheer number of possible variables that influence rates of non-persistence not included in the data collected make predictions of who will leave these programs problematic at best. The final classification tree model was able to predict, with between 74.1-75.9% accuracy, whether or not a student would persist in engineering education for three to four semesters after taking an introductory engineering graphics course for the test data set. Further, the model could accurately predict persistence between 89.5% and 93.8% of the time. However, the high incidence of Type I errors (false positives) indicates that the model does not well predict which students are likely to leave engineering programs, or the university itself, nor are the factors of non-persistence apparent in the model.

The classification tree model simply identified the factors most likely to predict persistence in the population studied. Further analysis of these variables and their statistical connection is necessary to understand why these variables may be

important and how this information can be used to increase persistence and, possibly, identify students who may be at risk of non-persistence. To best create a logical and consistent flow for the variable analysis, the individual variable discussion sections will follow the order they appear in the classification tree starting at the top. Due to the complexity and the existence of multiple models (with four or five variables), Figures 2a and 2b display the top three levels of both models for reference.

### *First-Generation College Students (FGCS)*

As previously noted, FGCS status represented a problematic variable in this model. Several reasons for this may exist. FGCS status is a complicated variable and represents a variety of potential issues and barriers faced by students whose parents did not attend college. FGCS are not a monolithic group and the barriers faced are affected by context, age, geography, and a myriad of other factors. Other issues such as definition, measurement, and perception add additional complexity to this variable that needs to be explored.

Exactly what FGCS status represents is not well defined in the literature. FGCS are most commonly defined as those students who neither parent has obtained a four-year degree. Enrollment and attendance in four-year programs and completion of two-year degree programs are not counted nor is near completion of a four-year degree program. Additionally, other family members (e.g. older siblings or grandparents) may have obtained degrees or are currently enrolled in college and provide supportive means similar to those of a parent with the credentials.

As with the definition of FGCS, measurement of this variable may also be problematic. FGCS status is self-reported in this study as it, along with many other demographic information, is protected data and held closely by the institution. Even if access was permitted by the university, FGCS status is still self-reported on the FAFSA applica-

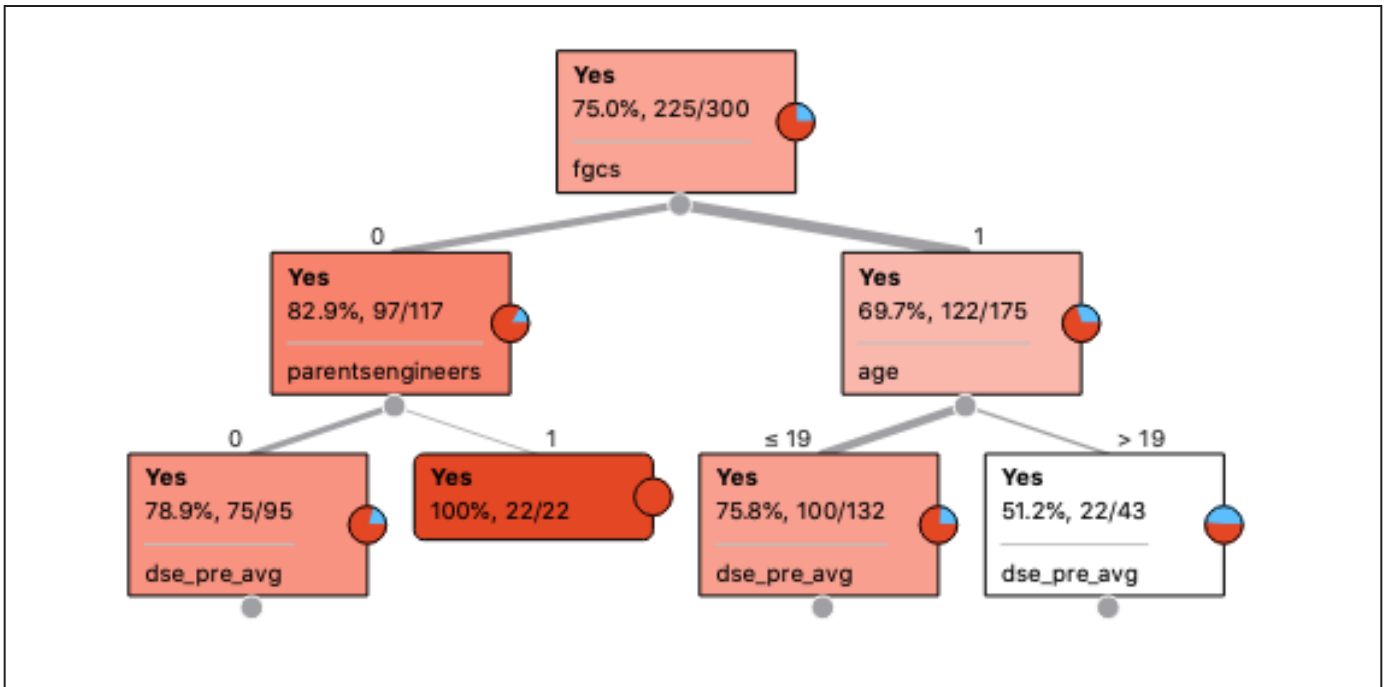


Figure 2a. Four variable model.

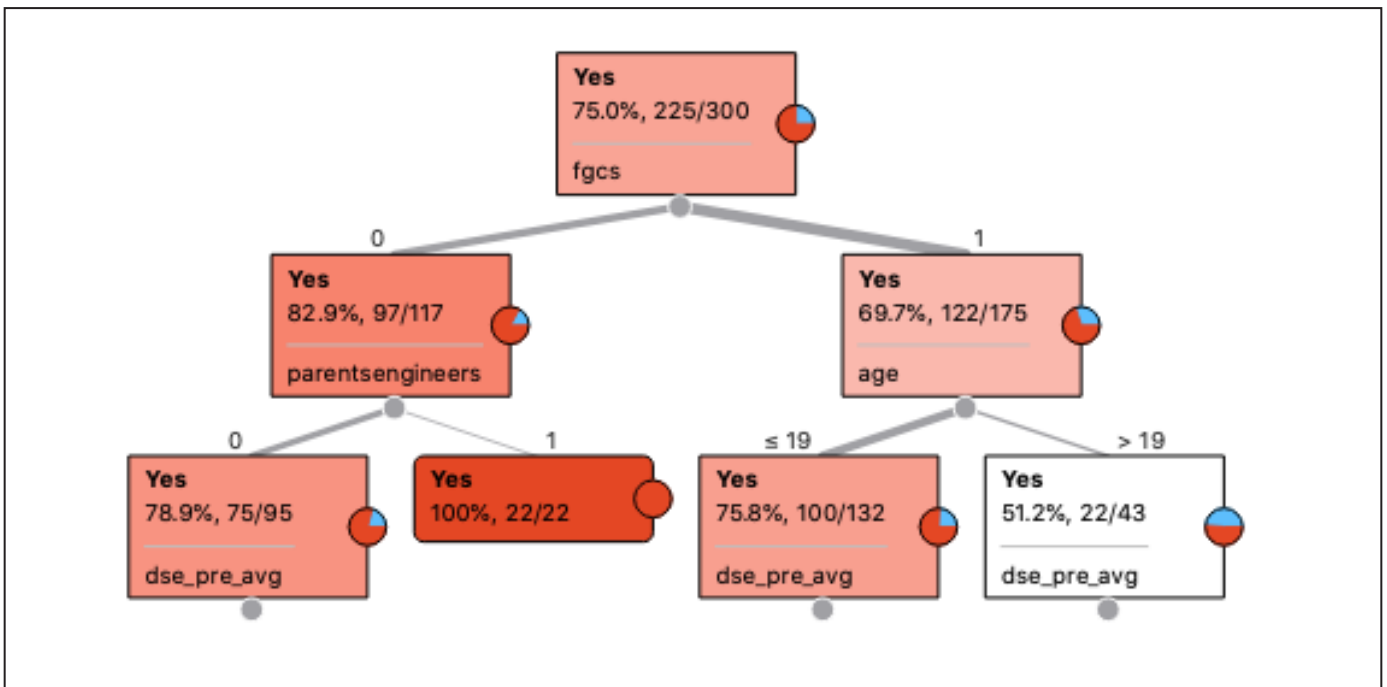


Figure 2b. Five variable model.

tion which is where the university gets the information. Because there is no verification of FGCS status, situations exist that could further confuse this variable such as adopted, fostered, and homeless students not having access to accurate information or selecting a set of “parents” who are not the ones who have the greatest influence on this variable.

A third concern also points to the previous issues related to definition and measurement, making FGCS status problematic. In this sample, 175 of 300 students identified themselves as FGCS. Among FGCS, 47 of them (26.8%) also stated that at least one of their parents were engineers with the underlying assumption being that to be an engineer, a four-year degree was required for licensing. Although it is possible to become an engineer without a degree, it is a difficult process and it is unlikely that more than a quarter of the students in this category would have a parent go through this process.

Whether the variables of FGCS status and whether a parent is an engineer are accurate is beyond the scope of this study and it is not possible to clarify with the studied population, it is a topic warranting future study. It also continues to emphasize that FGCS needs to be better understood in this context and how to address the needs of FGCS in engineering programs. Additionally, the definition of a parent may also be questionable in some cases. This too is outside the scope of this study but should be considered in this type of research in the future.

The personal experiences and perceptions of the lead author of this paper may shed some light on this issue. In this case, the lead author is not a FGCS, however, he would classify his father as an engineer even though his father never attended college and were it not for the education of his mother would be FGCS. The lead author has also been employed as an engineer, in two distinct fields, without a degree in engineering, although

it is in the related field of physics. Further, the issue of parental definition affects the lead author as he aged out of foster care and students in this category could consider foster parents’ professions and education levels, rather than birth parents, when answering one or both of these questions. While anecdotal, this scenario illustrates the possible confusion that exists when looking at the effect of variables such as FGCS.

Statistical tests alone cannot tell the entire story of FGCS status as nested groupings based on other variables have a significant impact on the probability of persistence. In an attempt to verify the results of the model, logistic regression was employed to ascertain the probability of persistence based on the FGCS variable. For this population, FGCS odds of persistence is 49.2% ( $OR = .492, p = .005, [95\% CI: .299, .808]$ ) based solely on this factor. The model clearly demonstrates that these other variables must be considered to work in conjunction with each other.

### ***Parents as Engineers***

Similar to FGCS, this variable is subject to perception and perspective. For this study, we simply asked, “Are either of your parents engineers?” The wording was intentional as the perception of whether their parents were engineers is as important as whether they have the credentials and, as this variable is also self-reported, the students may not know the specific degree or credential held by their parents. The perception of whether their parents are engineers or not engineers has an influence related to self-efficacy, motivation, and persistence.

This perception is important as described in the first author’s anecdotal experience as it may drive the expectations of the students. This is supported theoretically as a component of the Expectancy-Value Theory (EVT) of academic motivation (Parsons, Adler, & Kaczala, 1982). Under EVT, parents represent an environmental influence on their children’s self-perception of their academ-



ic abilities (Tomasetto, Mirisola, Galdi, & Cadinu, 2015). Whether a parent is actually an engineer may be irrelevant in that if a student believes them to be one, that perception becomes the reality, and that reality influences behavior even if not factually accurate. For this variable, in this study, the perception is what is important rather than a technical definition of an engineer.

Vicarious experience and social persuasion are major factors of self-efficacy that can directly relate to who a person knows who has a similar experience and whether those people can model behavior that supports the individual's goals (Bandura, 1997). These factors are especially important for women as self-efficacy in men is more strongly influenced by subject mastery than women (Denson et al., 2018). Having a parent as an engineer provides students with a person they can directly relate to and influence decisions related to the initial choice of engineering as a major and persistence within that major. For example, Kelly (2020) found that students who left engineering programs, at the same university in which this study was conducted, often described the experience as different from the expectations formed through high school "engineering" classes and descriptions of the work engineers perform. These differences were part of the reason students switched to a major outside of engineering. Students in that study viewed engineering as problem-based and hands-on whereas, in reality, college-level engineering education is largely math and theory with application coming later in their academic career if at all (Kelly, 2020). Having parents who understand these differences may provide students with a clearer picture of what engineering is before they matriculate and may provide the needed support to appropriately manage expectations.

Self-efficacy plays a key role in Social Cognitive Career Theory (SCCT; Lent, year) which also influences persistence in engineering. Specifically, SCCT constructs such as interest, outcome ex-

pectations, and social supports are influenced by a parent's experience and the support they can provide to the students. A parent being an engineer may also influence the student's choice of major and provide a more accurate description of the profession prior to entering an engineering education program. Shehab, Walden, and Wellborn (2015) found that interest and influence from others were the two largest contributing factors to choosing engineering as a major. Additionally, Mannon and Schreuders (2007) found that having engineers as parents strongly influenced the academic and career choice of engineering, especially among women. They also found that these family members may be passing on engineering-related knowledge, interests, and aspirations to children. This may then influence them to become engineering students and to remain in engineering education programs.

The presence of parental role models positively influences the persistence of the students in this study. For example, 100% of female students (19/19) whose parents are engineers were still enrolled in engineering three to four semesters later. Similarly, 100% of non-FGCS whose parents were engineers persisted (22 of the 22 students) with only nine of those students being female indicating that this variable affects both genders.

Similar to FGCS, whether a student has a parent who is an engineer adds no real value to our understanding of who will persist without an examination of other variables with which it interacts. For instance, logistic regression analysis on this variable alone provides no statistically significant effect between the variables ( $OR = 1.764, p = .079, [95\% CI: .937, 3.323]$ ). However, when FGCS is controlled for, logistic regression fails for non-FGCS students because it is a predictor (as noted previously with 22 of 22 students persisting). For FGCS, there is no statistically significant interaction ( $OR = 1.872, p = .119, [95\% CI: .850, 4.120]$ ).

Parents being engineers is a clearly important factor of persistence, especially among women and non-FGCS. Although issues with the FGCS variable exist, this result makes intuitive sense in that parents being engineers does not have as drastic an effect on FGCS as those students are less likely to have parents qualified for positions as engineers regardless of perception.

### Age

The age of a student in this study is the first variable that is not at the root node of any classification tree in the models discussed. The variable is most prominent when FGCS status is included in the model. As parents being engineers is the next most influential variable for those students who are not FGCS, age is the second variable that separates student groups who are FGCS. In both models (with and without FGCS), it is clear that the younger a student is when they take the introductory engineering graphics course, the greater the likelihood of persistence.

Remarkably, it is age and not class standing (i.e. freshman, sophomore, junior, and senior) that was the most relevant variable. In fact, there is no statistically significant association between age and class standing in this sample ( $r = -.064, p = .270$ ). A crosstabulation of these two variables makes the reasons for the lack of correlation apparent. Most of the students (49.7%) had freshman standing when they took the course, and most were aged 18 (35.5%) or 19 (37.8%). However, the distribution is not that linear, especially at age 20 and above with this population being evenly distributed across freshman, sophomore, and junior statuses. As this is an introductory course, there are fewer seniors taking it with only 5% of students being at this level; however, senior standing is largely held by 18 and 19-year-olds with only one-third being 20 or older. This trend may also explain the lack of correlation between age and class standing as many students come into universities (especially large, top-tier land-grants) with college credits from high school.

Traditional statistical analysis again fails to explain the influence of age on persistence. There is a weak and negative correlation between these two variables ( $r = -.160, p = .006$ ). Logistic regression does not bear any meaningful data as age is a continuous variable; however, simple linear regression supports the weak association with a small and negative coefficient ( $b = -.034, t(296) = -2.78, p = .006$ ). With a median age of 19, every year older represents a 3.4% percent additional likelihood of non-persistence with age only explaining 2.2% of the variance and only if the regression model does not include other variables ( $R^2_{adj} = .022, F(1, 294) = 7.71, p = .006$ ).

**FGCS.** In the classification tree model, age is most relevant for FGCS when that variable is included in the model. This model separates students based on age at 20-years-old. For FGCS 20 or older, only 51.2% persisted as opposed to 75.8% of FGCS 19 or younger. For non-FGCS, age is the last variable in one branch of the classification tree indicating it has the least influence on this population. It is only a factor for those non-FGCS whose parents are not engineers and have lower self-efficacy levels. Notably, this group is divided at 20 years old as well, with 63.2% of students 20 or older persisting contrasted with 91.3% of their younger peers persisting.

Age may be a factor for FGCS in that they may have fewer resources if they attend college as older students. Given the equal distribution across class status and the limited number of these students holding senior status (7.8%), it may be reasonable to assume these students entered the university later than their peers. It is also possible that their standing is lower due to failed courses which would then lead to a higher likelihood of non-persistence, but this analysis is not possible with the current data and is beyond the scope of this study. Regardless, more information is needed to determine the reasons age has an effect on persistence especially since it potentially runs counter to the assumption that older students



may be more mature than their younger peers. Given that it is most closely associated with FGCS status, the complexity of that variable, as discussed previously, may continue to affect all variables on that side of the classification tree.

**Non-FGCS.** When FGCS status is removed from the model, age becomes less impactful and the population affected is considerably narrower. Age is only included in the model for male students whose parents are not engineers. At this point, the model splits age into much more discreet groups with the first division occurring at 19 years of age. For those males whose parents are not engineers and are 18 or younger, 81.4% persisted. Those students over 18 at the time the course was taken only persisted 62.3% of the time and those over 20 persisted at a rate of 36.8%. Of those remaining students who were 20 years old or younger, only 53.6% of 20-year-olds persisted, further indicating age as a predictive factor for this specific sub-group. At 19 years or younger, self-efficacy is the next most influential factor as it was for male students whose parents were not engineers and took the course when 18 or younger.

The complexity of these figures is evidence of the reasons traditional analytic methods cannot accurately tell the story of how these variables interact and which sub-groups are most impacted by them. Neither the statistics nor the model can infer any causation between a student's age and whether they persist. What the model clearly indicates is that age is a factor for particular sub-groups. As these variables get further from the root node and the top of the classification tree, the more muddled their effect and the less reliable the statistics become. Further, a study into the reasons age is an important variable for particular groups is needed to better understand how to address the reasons students who take this class at a younger age persist at significantly higher rates. The impact is clear, but as FGCS status is a complex variable, so too is age since the

current information is limited, and only assumptions can be relied upon for this particular study and dataset to understand the true effect. Maturity, support, prior grades, and a host of other factors may be at play for this compound variable, and a better understanding of its underlying factors is needed to make recommendations based on age.

### **Gender**

In this study, female students persisted at significantly higher rates (85.5%) than male students (71.9%) with male non-persistence rates nearly double. This, as in the other variables, does not tell the whole story in that males also represented 77% of the population. Women have more barriers (both intrinsic and extrinsic) to entrance into engineering education which may explain the lower numbers and the higher persistence rates than men given that there is more to overcome prior to entrance and matriculation into engineering education programs. Barriers such as self-efficacy, lack of role models, and fewer pre-college educational engagement opportunities are established in the literature as barriers to entrance into engineering education (Hughes, 2012). This study does not support that being female is a risk factor for non-persistence in engineering education and demonstrates the opposite with markedly higher persistence rates in this study. The disparity in matriculation is another matter.

In the model that includes FGCS, gender does not appear in any node. Removing it slightly increases the classification accuracy of the model. The proportion of women who reported being FGCS is nearly equal to men (39.7% and 39.5% respectively) but the prominence of parents being engineers in that model may mask the relationship between gender and persistence. Both models have a subpopulation for which a parent being an engineer has a 100% persistence rate with the two groups being non-FGCS and women. Since nearly half of the non-FGCS group with 100%

persistence rates were women, the strength of having parents as engineers may be masking gender as all female students with parents as engineers persisted.

In the model without FGCS status, gender is the most prominent variable for both responses for parents are engineers. As stated earlier, all women with a parent who is an engineer persisted in this study. For male students with parents as engineers, 80% persisted with self-efficacy being the final factor for that population. For students whose parents were not engineers, self-efficacy was the most prominent factor of persistence for women and age, followed by self-efficacy, for men.

As with the other variables in these models, traditional statistical analysis fails to explain, or even detect, the effect of these factors as predictors of persistence. A correlation exists between gender and persistence ( $r = .133$ ,  $p = .022$ ), but this association is weak and provides little useful information. This, of course, changes when subgroups are controlled for but to properly identify these sub-groups without a classification schema is not only impractical but still fails to tell the story of how and why these factors interact and with whom.

### ***Self-Efficacy***

Specifically, this study examined the effect of three-dimensional modeling self-efficacy (3DSE) on persistence. Self-efficacy is defined as a person's confidence in their ability to assemble needed resources to complete a task successfully (Stajkovic & Luthans, 1998). Self-efficacy is domain-specific and is a well-established predictor of academic success (Zimmerman, 2000). In the context of the course and students used in this study, the 3DSE instrument (author, 2018) was used as three-dimensional modeling software is used heavily and represents an important area of mastery as not being able to effectively interact with the program is detrimental to course suc-

cess and continuation in engineering programs. The 3DSE assessment was given prior to the beginning of coursework and again after all coursework was completed. Both models discussed in this study found that the pre-assessment was the most important of the two in predicting persistence.

Unlike the other four factors discussed in this paper, self-efficacy is not a primary factor of persistence in the models. This is not to diminish the role of self-efficacy as it relates to persistence but to acknowledge that it is the most complex variable in this model both conceptually and statistically. It represents the only real continuous variable in the model—even though age is technically continuous, 90.2% of students were between 18 and 20—and is the only variable that is not strictly demographic in nature. Like the other variables, there is weak or no association between 3DSE and persistence ( $r = .015$ ,  $p = .796$ ) when examined alone. Self-efficacy appears most often in both classification tree models and is almost always the last node in the classification tree. This may be partially due to the continuous nature of the variable and that the models have differing split points depending on the sub-population in that branch of the classification tree.

There are some interesting revelations regarding 3DSE that have not been revealed in prior research. The relationship between 3DSE and persistence does not appear to be linear and, in some cases, a higher score decreases the likelihood of persistence. For example, FGCS under 20 years old persisted 96.7% of the time if their 3DSE score was between 60.63 (out of 100) and 77 but persisted only at 66.7% if the score was greater than 77. Similarly, for non-FGCS whose parents are not engineers, 83.6% persisted with scores below 70.5 but only 67.9% persisted with higher scores.

Female students whose parents are not engineers persisted 93.6% of the time with a 3DSE score



riers to engineering persistence. Engineering has longstanding gender-related factors that lead to female engineers not persisting in the field and historical gaps in interest and self-efficacy in girls prior to choosing a college major. This may have a direct impact on this particular source of self-efficacy that needs more examination in this context.

As mastery is an important source of self-efficacy, especially for males, this may partially explain why pre-course 3DSE being high may lead to attrition if those students do not perform academically at the same level as their expectations. This may support the possibility of overconfidence discussed previously and would benefit from further research that specifically investigated this phenomenon with the additional framework of expectancy-value theory included in the framework.

### Conclusion

What is most clear from the data is that these variables cannot be viewed in isolation and are both interconnected and interdependent; there is no single variable that can predict whether a student will remain in engineering education or not. It is also clear that more study into the factors of persistence and non-persistence is required and the true complexity of the variables and interactions between those variables needs to be analyzed. It is also clear from the data that persistence is better predicted than non-persistence. The number of potential reasons students leave academic programs or universities completely is too numerous to include in a single study and outside the scope of any individual paper.

Given the results of the classification and prediction models used in this study, the focus should be on the factors that lead to the successful completion of these programs rather than the reasons why students quit. Several of the factors discussed in this paper described well what

influenced student persistence at the university examined in this study over a one-year period. Although factors such as whether a parent is an engineer offer a more solid rationale for its influence on persistence, the other variables were not as clear. Factors subject to perception, definitional inconsistencies, and the non-binary or non-linear nature of FGCS and self-efficacy do not provide a clear reason why these variables influence persistence, however, the theoretical underpinnings of these constructs offer insight.

The models demonstrate these variables are important predictive factors among this population; however, it does not predict non-persistence. Further, there is not enough evidence that this model can predict persistence or identify students who will or will not continue in engineering education. This study was intended to identify the factors related to persistence for the purposes of improving courses, textbooks, program and college-level supports, and environmental conditions to increase the influence of these variables and potentially mimic the effects of those that cannot be addressed directly.

For example, it is not possible to make a student's parent an engineer or grant them a college degree with all of the experience that comes with those credentials and the influence passed onto their child. What can be done is to better understand how and why these variables influence persistence and incorporate methods and resources into engineering education in ways that provide the same, or similar, influences. Programs can incorporate more robust mentoring and outreach programs designed to address barriers students face during their academic careers that students who have parents with similar backgrounds and experiences would be able to receive. Programs could also create flexibility for older students who may have responsibilities uncommon to "traditional" college students such as children or job requirements.

Textbooks and course resources can incorporate imagery and materials that address the vicarious experience and social persuasion sources of self-efficacy. Affective factors can be addressed through relationship building with students and protocols developed to identify students in need of support. Expectations can be managed from at the beginning of courses and programs to address over or under-confidence that may lead to students leaving programs and exit interviews and surveys given to students who do leave to better understand the reasons for attrition and improve programming.

None of these potential solutions are novel and many have been done with varying levels of success and fidelity. Nonetheless, there remains a persistence issue in engineering education. The purpose of this study was to better identify the factors of persistence in engineering education and use this information to specifically target these factors through interventions. The current interventions may be too broad and offer generalized information and resources rather than target specific variables and the variable interactions that have the greatest influence on persistence.

The goal of this paper was to offer a foundation for future research and present a potential framework for intervention for improving persistence rates in engineering education. This study presents a unique model to investigate the factors that most influence student persistence in engineering education three to four semesters after taking an introductory engineering graphics course that includes three-dimensional modeling. Gender, self-efficacy, age, having a parent who is an engineer, and FGCS status emerged as the most influential predictors of persistence. The model created and analyzed in this research can predict persistence, among the population studied, with a high degree of accuracy but is limited in its ability to predict non-persistence. The use of machine learning and classification modeling

is an effective method to visually represent and identify sub-populations not apparent through traditional statistical analysis methods proved valuable and provided insight not available using other means. The lack of literature using this method in engineering education provides an opportunity for future research using these, and similar, methods.

### **Limitations and Future Research**

More research is needed to understand the interaction between these variables, identify other variables that may increase the accuracy of the model, or provide a clearer method of identification that needs to be explored. Additionally, this method should be evaluated with other populations and institutions. This study does not include all potential variables and could be expanded to include other factors including those that may be unique to specific universities and programs such as geographic and local economic concerns. This study included 3DSE, but the domain specificity of self-efficacy demands that other self-efficacy domains be examined. Although the sources of self-efficacy can be generalized as a lens through which to view other constructs of import from a theoretical perspective, the measurement of self-efficacy across differing engineering domains is another matter and needs to be explored further.

The exploratory nature of this study also presents a limitation. This data is not generalizable to the broader field of engineering education and needs to be further evaluated with a larger and more diverse population. This research is intended to propose a unique model for understanding persistence with the intent of spurring further research that expands the scope of this work and deepens our understanding of persistence in engineering education. It is also not meant to create an identification protocol by which to target resources to specific students or use in admission criteria. It would be impractical and un-



ethical to only admit students whose parents are engineers or have 3DSE scores within a particular range. Instead, this work is intended to provide a foundation for intervention that is applied broadly yet meant to influence particular populations of students. These interventions then need to be carefully examined against the persistence model used to determine its efficacy.

The caution against targeting interventions, or admission, for specific students or populations, is extremely important when focusing on non-persistence. The model can reliably predict students who will persist, but the number of false positives demonstrates that predicting non-persistence is problematic at best. The inclusion of a greater number of variables known to lead to student attrition that could increase the accuracy and an examination of those non-persisting students (especially those students with classification errors) are crucial next steps to this research. Further, more longitudinal information is needed to investigate these factors' true effect on persistence, graduation rates, and career longevity. The scope of this research project makes that impossible at this time and longitudinal research is generally difficult to perform. Regardless, this information is crucial to a more complete understanding of the effect of models and interventions on long-term persistence in engineering and engineering graphics.

### Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. REDACTED.

### References

Amin, S., Sanders, M. M., Kidd, A. E., & Rambo-Hernandez, K. E. (2023, June). How much does readiness matter? An examination of student persistence intention and engineering identity. *2023 ASEE Annual Conference & Exposition*. <https://doi.org/10.18260/1-2--43375>.

Besterfield-Sacre, M., Atman, C. J., & Shuman, L. J. (1997). Characteristics of freshman engineering students: Models for determining student attrition in engineering. *Journal of Engineering Education*, 86(2), 139-149. <https://doi.org/10.1002/j.2168-9830.1997.tb00277.x>.

Chang, C. N., Lin, S., Kwok, O. M., & Saw, G. K. (2023). Predicting STEM major choice: A machine learning classification and regression tree approach. *Journal for STEM Education Research*, 6(2), 358-374. <https://doi.org/10.1007/s41979-023-00099-5>.

Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data mining toolbox in python, *Journal of Machine Learning Research*, 14(1), 2349–2353. <https://dl.acm.org/doi/abs/10.5555/2567709.2567736>.

Denson, C. D., Kelly, D. P., & Clark, A. C. (2018). Developing an instrument to measure student self-efficacy as it relates to 3D modeling. *Engineering Graphics Design Journal*, 82(3), 1-9.

Ernst, J. V., & Clark, A. C. (2014). Self-regulated learning of at-risk engineering design graphics students. *Journal of Engineering Technology*, 31(2), 26-31. Retrieved from <https://proxying.lib.ncsu.edu/index.php?url=https://www.proquest.com/scholarly-journals/self-regulated-learning-at-risk-engineering/docview/1770928905/se-2>.

Flanigan, R.L. (2024). Engineering students quit, but retention tactics abound. *The American Society of Mechanical Engineers*. <https://www.asme.org/topics-resources/content/engineering-student-quit>

Hughes, R. (2012). Gender conception and the chilly road to female undergraduates' persistence in science and engineering fields. *Journal of Women and Minorities in*

- Science and Engineering*, 18(3), 215-234. <https://doi.org/10.1615/jwomenminor-scieng.2013003752>
- Kelly, D. L. (2020). Major Motivations: *First-Generation College Students' Motivations When Selecting Technology and Engineering Education as a Major*. North Carolina State University. <https://www.proquest.com/docview/2406470229?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses>
- Kelly, D. P. (2017). *Measurements of self-efficacy in engineering graphics students: An examination of factors impacting student outcomes in an introductory engineering graphics course* (Order No. 10758897). Available from Dissertations & Theses @ North Carolina State University @ Raleigh; ProQuest Central; ProQuest Dissertations & Theses Global; Social Science Premium Collection. (2020844328). Retrieved from <https://proxying.lib.ncsu.edu/index.php?url=https://www.proquest.com/dissertations-theses/measurements-self-efficacy-engineering-graphics/docview/2020844328/se-2>.
- Kelly, D.P., Clark, A.C., & Ernst, J.V. (2019). Online Active Learning Modules to Address Long-standing Gaps in Engineering Education for Students At-Risk of Non-Persistence. In J. Theo Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning* (pp. 1662-1669). Amsterdam, Netherlands: Association for the Advancement of Computing in Education (AACE). Retrieved October 22, 2025 from <https://www.learntechlib.org/primary/p/210189/>.
- Kodati, S. & Vivekanandam, R., (2018). Analysis of heart disease using in data mining tools Orange and Weka. *Global Journal of Computer Science and Technology: C Software & Data Engineering*. 18, 17-21. [https://global-journals.org/GJCST\\_Volume18/4-Analysis-of-Heart-Disease.pdf](https://global-journals.org/GJCST_Volume18/4-Analysis-of-Heart-Disease.pdf)
- Lee, Y. K., Freer, E., Robinson, K. A., Perez, T., Lira, A. K., Briedis, D., Walton, S.P. & Linnenbrink-Garcia, L. (2022). The multiplicative function of expectancy and value in predicting engineering students' choice, persistence, and performance. *Journal of Engineering Education*, 1(3), 531-553. <https://doi.org/10.1002/jee.20456>
- Lockhart, M. E., & Rambo-Hernandez, K. (2023). Investigating engineering identity development and stability amongst first-year engineering students: a person-centered approach. *European Journal of Engineering Education*, 49(3), 411-433. <https://doi.org/10.1080/03043797.2023.2262412>
- Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. *Research in Higher Education*, 59(3), 382-400. <https://doi.org/10.1007/s11162-017-9473-z>
- National Research Council. (2007). *Rising Above the Gathering Storm*. The National Academies Press. <https://nap.nationalacademies.org/catalog/11463/rising-above-the-gathering-storm-energizing-and-employing-america-for>
- Parsons, J. E., Adler, T. F., & Kaczala, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development*, 53(2), 310-321. <https://doi.org/10.2307/1128974>
- Shehab, R. L., Walden, S., & Wellborn, E. (2015). Motivating Factors for Choosing Engineering as Reported by Racial and Ethnic Minority Students. *2015 ASEE Annual Conference and Exposition Proceedings*, 1-26. <https://doi.org/10.18260/p.24507>
- Sorby, S. A., & Baartmans, B. J. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year engineering students. *Journal of Engineering Education*, 89(3), 301-307. <https://doi.org/10.1002/j.2168-9830.2000.tb00529.x>

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124(2), 240. <https://doi.org/10.1002/j.2168-9830.2000.tb00529.x>

Tomasetto, C., Mirisola, A., Galdi, S., & Cadinu, M. (2015). Parents' math-gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemporary Educational Psychology*, 42, 186-198. <https://doi.org/10.1016/j.cedpsych.2015.06.007>

Williams Jr, T. O., Ernst, J. V., Kelly, D. P., & Clark, A. C. (2019). Confirmatory factor analyses of the PSVT: R with data from engineering design graphics students. *The Engineering Design Graphics Journal*, 83.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-91. <https://doi.org/10.1006/ceps.1999.1016>

### About the Authors

**Dr. Daniel P. Kelly**, DTE, is an Assistant Professor of Technology, Design, and Engineering Education at NC State University. His research interests include increasing educational and career opportunities and access for students at high risk of non-completion of their academic programs. Dr. Kelly additionally focuses on youth in non-parental custody including foster care and justice-involved/incarcerated minors.

**Email:** [dpkelly@ncsu.edu](mailto:dpkelly@ncsu.edu)

**Dr. Erik J. Schettig** is an Assistant Teaching Professor in the Technology, Engineering, and Design Education program within the Department of STEM Education at NC State University. Drawing on more than 15 years of experience as both a K-12 educator and university instructor, he focuses on advancing technology and engineering teacher preparation and strengthening student success in STEM. An active member of professional organizations such as the International Technology and Engineering Educators Association (ITEEA) and

the American Society for Engineering Education, he has held leadership roles in the Engineering Design Graphics Division. Dr. Schettig's contributions have been recognized through awards for excellence in teaching, authorship, and professional service.

**Email:** [ejschett@ncsu.edu](mailto:ejschett@ncsu.edu)

**Dr. Aaron C. Clark**, DTE, is Co-coordinator of the Engineering Education program at NC State and Professor for Science, Technology, Engineering and Mathematics Education within the College of Education at North Carolina State University. He is a member of the Technology, Engineering and Design Education faculty. Research areas include graphics education, engineering education, visual science and professional development. He has also served in various leadership roles in disciplines related to engineering education and career and technical education. Dr. Clark is recognized as a Distinguished Technology Educator by the International Technology Engineering Education Association and for the American Society of Engineering Education; Engineering Design Graphics Division.

**Email:** [aclark@ncsu.edu](mailto:aclark@ncsu.edu)

**Dr. Jeremy V. Ernst** is Vice President for Research and Doctoral Programs and Professor of Technology at Embry Riddle Aeronautical University. He has held prior academic and administrative appointments at Virginia Tech and North Carolina State University. Dr. Ernst's research focuses on students identified as having non-matriculation indicators in STEM majors. His efforts center on providing evidence-based models that promote engagement, development of cognitive competency sets and performance-based application abilities.

**Email:** [Jeremy.Ernst@erau.edu](mailto:Jeremy.Ernst@erau.edu)