Retention Models for STEM Majors and Alignment to Community Colleges: A Review of the Literature

Jennifer Snyder
Valencia College

Elizabeth A. Cudney, Ph.D.
Missouri University of Science and Technology

Abstract

During the last decade, there have been numerous reports detailing the importance of increasing science, technology, engineering, and math (STEM) majors in the United States. Simultaneously, an increasing number of studies are being developed to predict a student's success and completion of a STEM degree, recognizing that retention is a significant issue for STEM majors. A majority of the studies focus on traditional college students that attend college directly after high school, which is no longer the model of the majority of college students. A growing number of students delay entry into college and do not enter through traditional routes. One of the growing entry points for STEM students is the community college or two-year institution. These institutions have grown in popularity due to tuition increases and lack of preparedness for traditional selective universities. As the need for more STEM majors and a diverse workforce increases, more research should be directed towards this growing pool of students. Retention models should investigate unique retention causation factors more thoroughly to address these STEM students and this pipeline. This research provides a systematic review of the literature on retention models for STEM education and provides a discussion of future opportunities to align predictive models with community colleges.

Keywords: Higher Education, STEM Education, Community College, Retention, Predictive Models

I. Introduction

After the worldwide economic downturn of 2008, there continues to be considerable apprehension and scrutiny surrounding the nation's economy and how to guard against weaknesses in the new global economy. There is strong evidence to support the assertion that Science, Technology, Engineering, and Math (STEM) careers will drive the economy of the future and help the United States remain globally competitive (Committee on Prospering in the Global Economy of the 21st, National Academy of Engineering Institute of, & National Academy of, 2007; Olsen, 2014; Vilorio, 2014). Further, students with substantial math and science training will experience more demand in the workforce, even if not working directly in STEM careers, due to enhanced critical thinking skills (Council et al., 2013). Data from the Bureau of Labor Statistics shows employment in STEM fields is expected to increase by approximately one million jobs between 2012 and 2022 (Vilorio, 2014). In light of these growing concerns, former President Obama challenged the country to increase the number of STEM graduates by one million in this ten-year period (Olsen, 2014). In response to his call, the President's Council of Advisors on Science and Technology (PCAST) organized a report on the strategies that could help attain this goal. In *Engage to Excel*, PCAST addressed the important points of retention, community colleges, and the need for more diversity, which this review of the literature will investigate more deeply (*Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, 2012). Despite intensified efforts, the U.S. has seen a decrease or stagnation in the number of STEM majors in recent decades (Snyder, Dillow, & Staff, 2012).

While there is some scrutiny about the heterogeneity within the STEM labor market, there is little argument on the need for more diversity in these fields (Committee on Underrepresented et al., 2011; Terenzini, Lattuca, Ro, & Knight, 2014; Xue & Larson, 2015). The engineering workforce should mirror the diversity of our population if it is going to keep pace with the global markets (Hagedorn & Purnamasari, 2012; Starobin & Laanan, 2008; Terenzini, Lattuca, Ro, & Knight, 2014; Xue & Larson, 2015). In 2015, *Solving the Equation: The Variables for Women's Success in Engineering and Computing* illuminated the gender inequity in STEM degrees, especially engineering and computing. These two segments of STEM account for 80% of the workforce, yet women are profoundly underrepresented. Women account for a minor fraction of the engineering and computing workforce, representing just 12% and 26%, respectively. The numbers are more drastic when one considers women of color (Committee on Underrepresented et al., 2011; Costello, 2012; Dika & D'Amico, 2016). Increasing access for women to STEM careers is proposed to help close the gender wage gap (Costello, 2012).

Recent data from governmental sources makes a compelling argument for attention to STEM majors and retention.

- Students are choosing STEM majors in sufficient numbers as a whole with approximately 28% of bachelor's degree students and 20% of associate's degree students choosing a STEM major at some point within six years of entering higher education (Chen, 2013).
- Rates of U.S. undergraduates that choose STEM majors trail key competitors and the number has not increased drastically in decades (Chen, 2013).
- The percent of women enrolled in science and engineering has remained flat from 2000-2013 (National Science Board, 2016).
- Of U.S. citizen and permanent resident science and engineering doctorate recipients, 18.4% reported earning college credit from a community college with the percent ranging from 12.7% for Asian to 32.3% for American Indian ethnicity (National Science Board, 2016).
- Of students receiving a bachelor's degree in science and engineering, 18% had previously earned an associate's degree (National Science Board, 2016).
- Of the associate degree STEM entrants, 69% left the fields. Of these, 43% of female associate's degree students switched out of STEM, while only 29% of their male peers left (Chen, 2013).

When looking at the national goal of increasing STEM majors, there must be a thorough analysis of retention (Drew, 2011; Seymour & Hewitt, 1997). PCAST recommended efforts be guided toward increasing the retention of students, since a minor increase in retention could have significant benefits in the total number of graduates. STEM retention is currently reported to be 48% nationally, which is an average of all reporting institutions (Chen, 2013). The numbers are more telling when looking at institutions as sectors. Science and Technology institutions have much higher retention due to various factors and rigorous admittance requirements. The lowest retention (30%) of STEM majors is seen within community colleges, which struggle with open enrollment and lack of academic preparedness in many students (Chen, 2013).

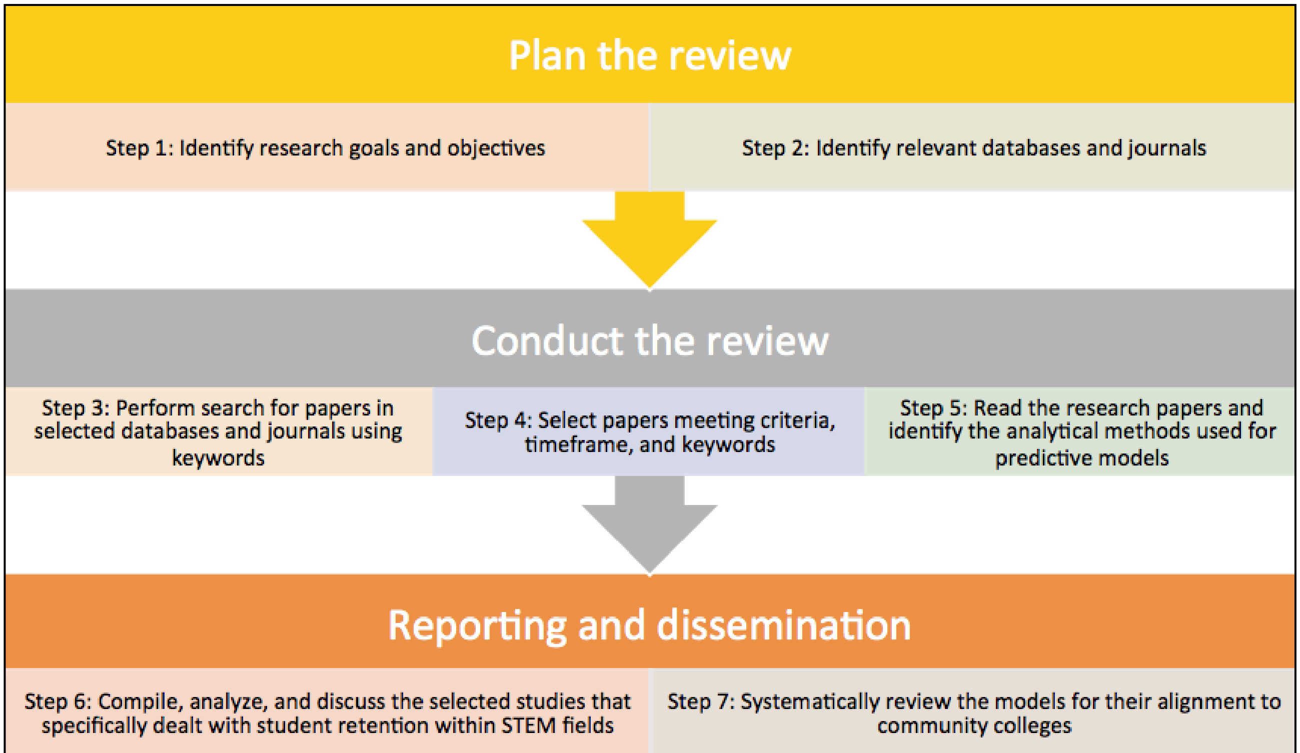

Figure 1. Research methodology for systematic literature review

Retention increases could help achieve the goals set forth by former President Obama and allow the U.S. to remain competitive in this increasingly important segment of the economy.

One population often overlooked in tackling the nation's goal for increasing and diversifying STEM graduates is the community college transfer student (Hoffman, Starobin, Laanan, & Rivera, 2010). In multiple National Science Foundation (NSF) reports, there is growing evidence that community colleges are critical to increasing the diversity of STEM, especially in engineering (Committee on Underrepresented et al., 2011; Hagedorn & Purnamasari, 2012; S. Starobin & Laanan, 2008). In *America's Overlooked Engineers*, data outlines that community colleges have a much more diverse student population pursuing engineering degrees. However, when studying engineering graduates there is little difference in ability between graduates that attended a community college and those that received all credit from a four-year institution (Terenzini et al., 2014).

Community colleges currently educate almost half of the country's undergraduate students, including STEM majors (Hagedorn & Purnamasari, 2012; Starobin & Laanan, 2010). Additionally, the community college student population is much more diverse than universities due to flexible schedules, open enrollment, and lower tuition (Cohen, Kisker, & Brawer, 2014; Jackson & Laanan, 2011). In light of these factors, the community college system should be a major partner and contributor to the STEM degree pathway. As a research community, there should be more investigation into this overlooked resource for quality, diverse undergraduate transfer students. Given that community colleges have the lowest retention rates, it is important to remember that most students leave STEM within the first two years (Chang, Eagan, Lin, & Hurtado, 2011; Seymour & Hewitt, 1997). Increasing community college retention rates could have a drastic impact on the average STEM graduation rates while also potentially diversifying the workforce. Ultimately, there cannot be substantial changes to retention rates without working with community colleges, yet little academic research is focused on this sector of higher education.

Higher Education Institutions (HEIs) must develop clear strategies to recruit and retain STEM majors to assist in the national effort to produce quality students. This paper will outline the importance of STEM majors, the significance of retention values in maximizing our country's economic competitiveness, survey existing predictive models, and highlight the growing need to incorporate community colleges in the national dialogue.

The remainder of this paper will be broken into sections. Part II will provide the literature review methodology. Part III will review the various retention causation factors and predictive models currently being used by colleges and universities and highlight the reliability of models and development methods employed. Part IV will relate the retention factors and models to community colleges and show how the current models do not address a majority of community college students. Part V will highlight opportunities to modify these models to properly address community college students.

II. Research Methodology

The purpose of this systematic literature review was to examine current literature relating to the use of predictive models in STEM retention, specifically in community colleges. The research results were compiled and analyzed according to the methodology introduced by Tranfield, Denyer, and Smart (2003). The research was conducted per the flow of processes shown in Figure 1.

Plan the review

The review was limited to Google Scholar, Education Resources Information Center (ERIC), Web of Science, IEEE, and SCOPUS. Additionally, there was a search of the *Journal of Engineering Education, Community College Journal of Research and Practice, Community College Review, and ASEE Journal of Engineering Technology*. A thorough search for "student retention" and "STEM" and "community college(s)" and "predictive models" did not yield any results in the chosen databases. With the lack of published research pertaining to community colleges hindering the results, the review was expanded by excluding the term "community college(s)" in the search factors. Recognizing the use of predictive analytics is ever evolving, the search was limited to the timeframe of the year 2000 to the present. The keywords searched were manipulated to attempt a larger review pool given the synonymous use of the terms retention and persistence. While the two terms represent different concepts, they are used interchangeably in the literature. The search criteria included a combination of the following keywords: "STEM or science or engineering" and "student retention or persistence" and "predictive model". The search of community college specific journals did not yield as many results as suspected and few articles developed a retention predictive model specifically targeting STEM students.

Conducting the review

The search for relevant papers did not yield many results. The most robust search was within the Journal of Engineering Education for the keywords "persistence" and "predictive model", which returned thirty-four articles. Those articles ranged in predictive models from career choice to persistence in a specific course. Several studies provided retention models that were developed to predict the retention of students based on various causation factors. There is increasing interest in data analytics being used to aid retention as presented in Figure 2, which highlights the number of articles found by year.

The number of articles returned in the searches was misleading in many cases due to "student retention" being a keyword with multiple meanings. For this analysis, the focus was on STEM retention from freshman year through graduation.

III. Retention Causation Factors and Current Predictive Models

Retention of college students has been a focus of research for decades. There is substantial belief that increasing the retention of students will benefit every sector of our country (DeBerard, Spelmans, & Julka, 2004; Li, Swaminathan, & Tang, 2009). In order to directly impact retention rates, it is necessary to understand the causation factors that impact the persistence and completion rates of students.

Emerging in the last half of the twentieth century were two seminal pieces of research on retention and the factors that contribute to attrition. Tinto (1987) and Astin (1993) produced significant research on retention and contributing factors. Both studies investigated student attributes, but also the institutional effects that influence a student and their decision to complete college or withdraw.

In *Leaving College*, Tinto describes in depth the various causation factors that lead a student to withdraw. Tinto's model examined individual and institutional factors that led to a student's decision to voluntarily withdrawal (Tinto, 1987). The individual factors of intention and commitment seem to be critical attributes lending to a student's success in college. These are qualities that a student has before entering college and can be influenced, but these qualities are individual in nature. Institutional factors are the variables that can be impactful after a student enters the higher education system. These factors speak to the student's overall integration into the institution. The factors are adjustment, difficulty, incongruence, and isolation. One of the most significant relationships appears to be between a student and faculty. It should be recognized that a negative interaction with faculty or staff can lead a student to feel less connected to the institution and influence their decision to withdraw. Tinto highlighted the importance of understanding institutions as systems and viewing the model from a longitudinal perspective with interacting components (Tinto, 1987).

In *What Matters in College*, Astin also studied retention and factors that influenced it. The model Astin produced is referred to as the I-E-O model. It emphasizes the importance of the input (I) to the system, which is the background and preparation a student brings to the institution. The institutional environment (E) has an effect on those inputs and together will determine the outcome (O). This study also emphasizes the engagement of students with the institution (Astin, 1993).

Using these models as a springboard, Seymour and Hewitt (1997) focused on STEM majors in the book, *Talking About Leaving*. The overall aim of this research was to identify sources of qualitative differences in students' experiences when pursuing a science, math, or engineering (SME) degree. The research investigated what institutions and departments did that encouraged attrition amongst the SME majors, while also comparing the attrition causes of females and minorities to that of the majority. One of the largest findings is that there was not a significant difference in cognitive ability between "switchers" and "stayers". The four most common factors of switching were loss or lack of interest in science, non-SME degrees held better educational opportunities, poor teaching by SME faculty, and feeling overwhelmed by the pace and load of a SME curriculum. When questioning students, it was found that the weed out curriculum of SME degrees is a factor in their feelings of being overwhelmed. Students felt faculty did not understand that the weed out system favors students that are independently funded. This is problematic given the need to diversify SME and increase success of students from lower socioeconomic backgrounds. When exploring the gender differences in SME retention, it was found that women were more likely to choose their degree due to an active influence of others. Females also reported feeling alienated in their programs, which possibly leads to the higher attrition rate seen for female SME majors. Further, poor high school preparation was claimed by students of color and women more than other classes of students. Overall, the causes of high attrition rates amongst these majors was as variable as Tinto and Astin found for all majors; however, it does appear that SME majors suffer more from a weed out mentality of faculty and poor teaching (Seymour & Hewitt, 1997).

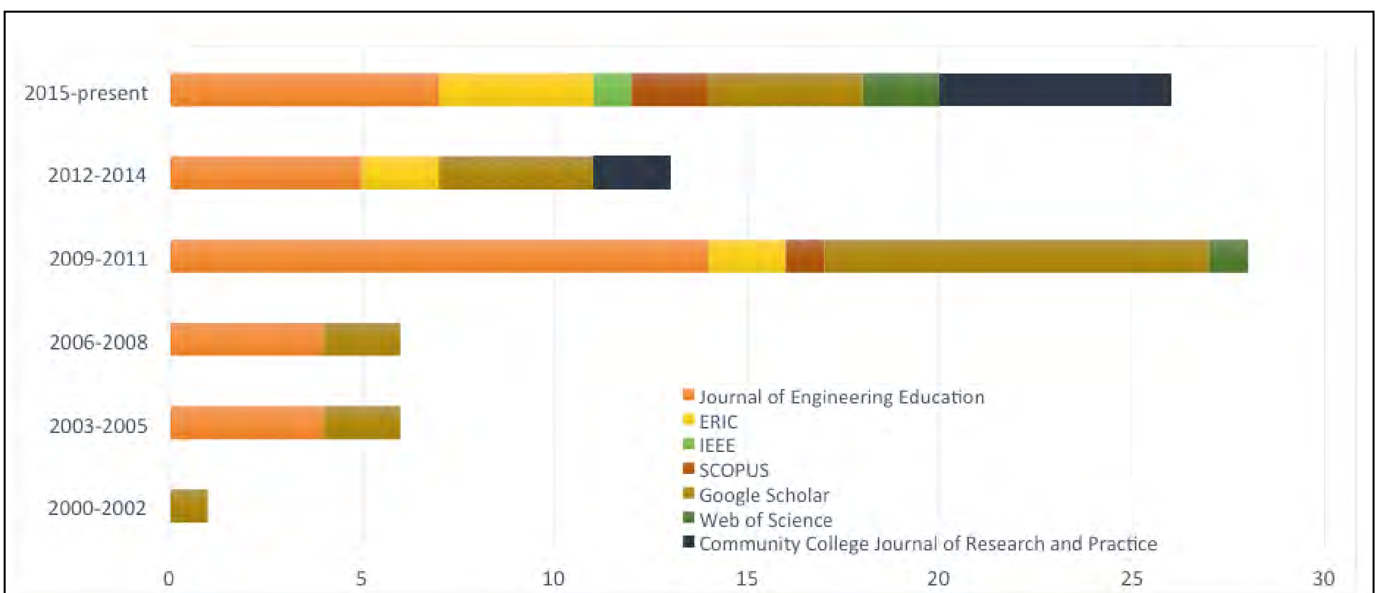

Figure 2. Publications by year and database/journal searched

Study	Description	Method	Key Findings
Burtner (2005)	Studied non-cognitive factors and their impact on student retention in engineering curriculum compared to those that left the college or university.	Discriminant Analysis, ANOVA, Regression Analysis	• Confidence in STEM • Intrinsic motivation
French et al. (2005)	Examined persistence and achievement of engineering students investigating cognitive and non-cognitive factors relating to the predictive worth of variables.	Regression Analysis	• High school GPA • High school rank • SAT math • Motivation
Bernold et al. (2007)	Investigated the learning styles of students and the impact it had on their retention in engineering curriculum given the traditional lecture model of engineering education.	The research correlated student outcomes to learning styles as determined by the Learning Type Measure (LTM).	• Learning styles could predict student success in engineering
Nicholls et al. (2007)	Investigated the variables that can predict a student's intention to major in STEM versus non-STEM based on quantitative and qualitative indicators.	ANOVA, Regression Analysis	• SAT math • High school GPA • Self-reporting math, computer, and academic ability
Veenstra, Dey, and Herrin (2008)	Explored the differences in predicting success for engineering versus non-engineering students to detect any significant differences.	Factor Analysis, ANOVA, Regression Analysis	• Higher ACT math and science scores • Higher self-ratings in math and computers
Moses et al. (2011)	Investigated the causation factors that contributed to freshman year retention of students in an engineering program.	Regression Analysis	• High school GPA • ALEKS score • Openness subscale of NEO-FFI
Marra, Rodgers, Shen, and Bogue (2012)	Analyzed students that left engineering and investigated what factors influence student retention and how those factors differed according to gender.	Exploratory Factor Analysis, Regression Analysis	• Poor teaching • Curriculum • Feelings of lack of belonging
Alkhasawneh and Hargraves (2014)	Developed a hybrid model using machine learning and qualitative surveys to predict retention of underrepresented minorities.	Neural Network	• High school math and science • Race • Gender • Freshman year grades
Hall et al. (2015)	Investigated first-year students using quantitative factors and Neuroticism-Extraversion-Openness Five-Factor Inventory (NEO-FFI) to develop a model for predicting retention for persisting students versus those that left engineering.	ANOVA, Regression Analysis	• High school GPA • ALEKS score • Conscientiousness
Morganson et al. (2015)	Employed the Embeddedness Theory to determine factors that cause students to persist in STEM by specifically looking at the reasons students stay.	This research employed the consensual qualitative research (CQR) approach, which included open-ended questions. The answers were analyzed in a structured format.	• Fit emerged as a significant aspect • Identification with major was found more important than with the institution

Table 1. Research contributions in STEM student retention

There are still several variables not understood in student decision making about withdrawing from an institution, but what is clear from the research is the causes do not lie squarely on the individual student. There seems to be a relationship between a student's individual characteristics and their experiences with the institution. Following these seminal research studies on retention, there have been multiple recent studies into the student and institutional factors that can predict student success in STEM. There are several causation factors that appear relevant in these retention studies. Most studies concentrate on the quantitative factors a student possesses before entering higher education such as high school GPA, high school rank, and standardized exam scores. Recognizing the complexity of the causation factors, studies usually include a multifaceted approach to the investigation including both quantitative and qualitative variables.

Several studies examined the combination of qualitative and quantitative factors and found student motivation and confidence significantly impacted their success and retention (Burtner, 2005; Eris et al., 2010; French, Immekus, & Oakes, 2005; Nicholls, Wolfe, Besterfield-Sacre, Shuman, & Larpkiattaworn, 2007). Morganson et al. (2015) investigated a different view of retention by studying the factors that influence a student to stay and complete a degree using the Embeddedness Theory. The Embeddedness Theory looks at fit, link, and sacrifice to determine factors that anchor a student to their degree and institution (Morganson, Major, Streets, Litano, & Myers, 2015). Bernold et al. (2007) studied learning styles and the influence they had on retention and success (Bernold, Spurlin, & Anson, 2007). The study shows that learning styles most conducive to the traditional lecture pedagogy of engineering curriculum have a higher retention rate. From a gender perspective, females were more likely to have a learning style that did not perform well in the traditional engineering lecture style (Bernold et al., 2007). Table 1 summarizes the various important contributions to the study of retention regarding STEM students.

It is clear from studies there is importance in a student's cognitive and non-cognitive abilities relating to the prediction of success and persistence. These studies reinforce Seymour and Hewitt's (1997) findings on several of the causation factors relating to retention, but many researchers did not investigate the institutional factors that could provide a more reliable model to investigate both student factors and institutional contributions.

With a national goal of increasing retention in STEM majors, there have been several predictive models developed to help institutions target factors that could lead to increased retention. These models help institutions allocate budgets properly and plan for programs that enhance student completion. The studies in Table 1 used a variety of analyses to develop predictive models. Analytical methods were chosen based on the purpose of the research and the types of variables available. The common methods are highlighted next.

Regression analysis. Many of the studies highlighted in Table 1 used regression analysis in some form, as it allows for a complete analysis of factors and development of a model. Regression analysis is often used with historical data and can be useful in expressing relationships between predictive variables and a response variable (Montgomery, Vining, & Peck, 2012). Many of the studies in Table 1 used regression analysis to develop predictive models. In Marra et al. (2012), the study determined three factors were important to student retention: poor teaching and advising, curriculum difficulties, and lack of belonging. The analysis used simple linear regression and found the number of months students stayed in engineering was related to the predictive factors of poor teaching and advising and curriculum difficulties. The research also employed regression analysis to determine the predictive power of original confidence. A negative relationship was found between original confidence and the lack of belonging as a factor in retention. Multiple regression analysis was used to examine the impact of poor teaching and advising, curriculum difficulty, and lack of belonging on students' cumulative GPA. It was determined those three variables account for 20.7% of the GPA variation (Marra, Rodgers, Shen, & Bogue, 2012).

Veenstra et al. (2008) investigated the differences in modeling engineering versus non-engineering student success. Stepwise regression was used to determine the set of predictors for first year success for both engineering and non-engineering students. The results indicated that 37-38% of the variation of engineering students' first-year GPA was explained by pre-college characteristics, which were largely associated with academic preparation (Veenstra, Dey, & Herrin, 2008). French et al. (2005) investigated the cognitive and non-cognitive variables that were predictive factors for student success and persistence within engineering. Three regression analyses were performed using historical data collected from two cohorts of engineering undergraduate students. A hierarchical linear regression was used for predicting GPA and it was determined that several cognitive factors account for 18% of the variance. When predicting persistence in the university, only GPA was a significant predictive variable, which resulted in correct classification 89% of the time. The hierarchical logistic regression model for engineering students found more significant variables including GPA, high school rank, SAT-math, and motivation. This predictive model had correct classification 65% of the time (French et al., 2005).

Hall et al. (2015) found only one significant parameter for comparing persisting students with those that left in good standing. The odds of persisting increased by 1.63 for every one standard deviation on the SAT-math score. When comparing persisting students with those that leave in poor standing, there were three significant predictors including high school GPA, conscientiousness, and Assessment and Learning in Knowledge Spaces (ALEKS) score. The success of prediction depended on the group of students being analyzed, with persisting students (69.9%), left in poor standing (64.7%), and left in good standing (40.0%) varying in accuracy of prediction (Hall et al., 2015). DeBerard et al. (2004) successfully used regression analysis to predict GPA, but did not find statistically significant variables for predicting retention. This reinforces the multifaceted causation factors that likely exist for retention prediction.

Exploratory factor analysis. It is common to have a large set of data and use exploratory factor analysis to estimate the strength and direction of the influence of factors on a response. Exploratory factor analysis is a methodology to analyze data and explore significant factors, which allows for a predictive function of the exploratory factor analysis (Fabrigar & Wegener, 2012; Osborne, 2016). This technique is useful when there is not a suitable hypothesis and investigation of the data is warranted; such as when Marra et al. (2012) used exploratory factor analysis to determine which factors influence a students' decision to transfer out of engineering. The analysis identified five factors, with the first three factors explaining 65.92% of the total variance. The three factors were poor teaching and advising, difficult curriculum, and lack of belonging. Once those factors were identified, Marra et al. used regression as described previously (Marra et al., 2012). Li et al. (2008) used exploratory factor analysis to determine different perspectives students hold about engineering and generated four factors from the data with the interest factor being significant between engineering and non-engineering students (Li, McCoach, Swaminathan, & Tang, 2008). Many studies use exploratory factor analysis to isolate the factors required for further investigation with predictive modeling.

Machine learning. Machine learning has gained popularity as a method that might have the ability to increase the accuracy of predictive models in retention since it encompasses several techniques such as artificial neural networks (ANN) and decision trees. Decision trees use splits to generate a model and produce rule sets (Luan, 2002). Decision trees and neural networks offer advantages in predicting key outcomes over traditional statistics and have been shown to accurately predict students that would graduate within three years or less (Herzog, 2006).

Alkhasawneh and Hargraves (2014) used machine learning techniques and surveys to develop a hybrid model to predict first year retention in STEM. The study investigated underrepresented minority (URM) students compared to majority students. The model is a hybrid due to the inclusion of a qualitative survey given to a focus group attending a summer program. The neural network technique used FeedForward back propagation. The resulting hybrid model had an accuracy of prediction of 66% for URM, which was the lowest accuracy for the groups. The highest accuracy was found with majority students (Alkhasawneh & Hargraves, 2014). Djulovic and Li (2013) compared four techniques including Bayes model, C4.5 decision trees, neural networks, and rule induction with regards to their accuracy of prediction. All four techniques performed very well for predicting retention. The accuracy improved as more variables were added with a final accuracy of 98.81% for retained students using the rule inductive model (Djulovic & Li, 2013). Delen (2010) also found decision trees to be promising for accurately predicting students that will be retained. Regardless of the technique, there was a lack of sufficient accuracy in predicting attrition.

All of these methods have promise as tools to develop predictive models, but clearly more powerful methods should be investigated for use in community colleges. This is an area that is often overlooked in the development of retention models by researchers (Cohen, 2005).

IV. Retention Factors and Models in Community Colleges

As college tuition increases and completion time expands, community colleges have emerged as a viable option for students. Community colleges have been discussed heavily in politics lately as an important sector of higher education and their importance in keeping costs low while impacting the economy with workforce development (Swanger, 2013). Community colleges grew out of a democratic mission to offer post-secondary education to everyone (Cohen et al., 2014; Young, 1997) by offering many smaller communities both general education and technical job training. Community colleges remain close to their original mission of serving the local community with over 50% of community colleges being located in rural settings (Swanger, 2013). Since 1901, the establishment of the first community college, the mission has expanded and is seen as a comprehensive concept. One important aspect of community colleges is the concept of "open access" with an emphasis on developmental education and preparing students for transfer to universities (Cohen et al., 2014; Deegan, 1985; Hoffman et al., 2010; Swanger, 2013).

Community colleges serve a very diverse student population (Hoffman et al., 2010; Horn & Nevill, 2006). This diversity extends to the institutions themselves. Community colleges can be private or public, focus on transfer preparation or workforce development, and offer only associate degrees or select bachelor degrees. The academic and institutional diversity could contribute to difficulties in studying them (Hoffman et al., 2010).

When investigating women in community college, it is noted that a majority of community college students are female reaching approximately 58% of the student population (Hoffman et al., 2010). Costello (2012) reports that 20% of community college students are women with children and one in ten female students is a single mom. Even with this large population of females, the number of females pursing STEM degrees remains small (Hoffman et al., 2010; Packard, Gagnon, LaBelle, Jeffers, & Lynn, 2011).

Community colleges are much more racially congruent with the area in which they are located than most universities (Cohen et al., 2014; Hoffman et al., 2010). Additionally, 38.5% of community college students are racial minorities with Hispanic students representing the fastest growing sector (Hoffman et al., 2010). Unfortunately, data indicates that participation in STEM degrees is low for these demographics (Hoffman et al., 2010). Tsapogas (2004) noted that Hispanic Science and Engineering (S&E) graduates were more likely to have attended a community college, with approximately 51% attending before transferring to receive a bachelor's degree. Community colleges are a strong resource for diversifying STEM while providing the increasingly necessary preparation.

There are other factors that contribute to a more diverse demographic profile of community college students. Studies show 79% of community college students have jobs and work an average of 32 hours a week, which lends to more part time enrollment (Costello, 2012; Horn & Nevill, 2006). Data indicates that delayed entrants to college are more likely to favor a two year institution and this trend was especially noticeable when looking at minorities and women (Cohen et al., 2014). First generational college students (FGCS) are also more likely to begin their

Study	Description	Method of Analysis and Prediction	Key Findings
Armstrong (2000)	Investigated the predictive validity of placement exam scores on grades and retention in math and English courses at a community college to answer three research questions. (1) Are placement exams predictive of course outcomes? (2) Do student characteristics affect the prediction of course outcomes? (3) Do teacher characteristics affect the prediction of course outcomes?	Regression analysis was used to predict final grade given a large set of variables such as test score, demographics, dispositional and situational characteristics, and instructor data.	Research questions (1) The correlations coefficients between placements scores and grades failed to meet the 0.35 level for statistical validation. (2) Student variables tended to contribute significantly to the predictive model. It was found that previous performance in school was a strong predictor of success. • Part-time instructors had more variation in their grades, but overall the instructor data being added to the model only increased its validity.
Calcagno, Bailey, Jenkins, Kienzl, and Leinbach (2008)	Analyzed institutional factors to determine how the factors that correlate to outcomes are measured by student completion and transferring.	Examined institutional factors such as leadership, faculty relations, and local politics to determine the best fit of four models from binary outcome. Model 1 assumes the students' probability of success is only affected by observable institution factors of the first community college attended. Model 2 incorporates the institution's unobservable factors. Model 3 weights the multiple community colleges the student might be attending prior to the outcome. Model 4 uses the continuous factor of credits accumulated by the student.	• Size of institution, diversity of student population, and percent of adjunct faculty were found to have a negative impact on outcomes. • Found student completion was more closely correlated to individual factors instead of institutional factors.
Fike and Fike (2008)	Examined the Fall-to-Fall and Fall-to-Spring retention of first time in college (FTIC) students to determine which factors can be considered predictors of student success and retention.	Descriptive statistical study of retrospective data from an urban Texas community college leading to a logistic regression analysis for predictive modeling.	• Fall-to-Spring and Fall-to-Fall findings were very similar o Successful completion of a developmental reading course had a strong positive correlation with retention and persistence. o Successful completion of a developmental math class, receiving financial aid, taking online courses, and seeking student support services also had a positive correlation. o Student age and number of credits dropped first semester were negatively correlated. o Multivariate model explained 31% of variance.

Table 2. Community college retention factors and models (continued on next page)

Wai-ling Packard, Gagnon, and Senas (2012)	Evaluated the delays experienced by 172 STEM students in the transferring process from a two-year to four-year institution.	Student surveys from three community colleges in Massachusetts were used in a phenomenological study.	• Delays were attributed to poor advising, improper course alignment with transfer institutions, and lack of college resources. • Interesting finding was even with delays. Students report overwhelmingly positive feelings regarding the community college.
Wang (2013)	Investigated attributes influencing a student's decision to pursue a STEM degree and identify possible intervention predictors.	Social cognitive career theory and multi-group structural equation confirmatory factor analysis.	• Self-efficacy in math and interest in STEM were important attributes to all students. • Being married was positively associated with choosing STEM. • The number of remedial courses required had a negative impact on student decision to pursue STEM.
Mertes and Hoover (2014)	Examined the potential predictive variables for Fall-to-Fall retention at a community college. The scope included high school GPA, age, gender, ethnicity, credit hour load, educational goal, remedial need, and receipt of financial aid as factors that had been identified as significant in prior research.	Data was collected for a Fall 2007 and a Fall 2010 group and analyzed using Chi-square analysis to determine significance of prediction. Logistic regression was used to identify the combination of important factors that would best yield prediction of retention.	• The Fall 2007 and 2010 groups both showed significant Chi-square scores for age, gender, program of study, and CIS 100 grade, but Fall 2010 group also showed ethnicity, credit hour load, math placement score, and receipt of financial aid. • The regression results were incomplete due to missing data for a majority of students, but the results of available data showed predictive power for CIS 100, age, and program of study for the Fall 2007 group. The Fall 2010 group only showed CIS 100 grade and age.
Luke, Redekop, and Burgin (2015)	Investigated the relationship between career decision self-efficacy, career locus of control, student's connection between education and employment, and the intent to remain in school and complete it at a community college.	A survey was used to measure the variables of interest. The data was then analyzed using regression analysis to answer the research question pertaining to psychological impacts on intention to return to school and actual retention.	• Total self-efficacy and internal locus of control were negatively related to intent to return, while school/work connections were positively related to intent. Intention to return and internal locus of control were also predictive for students returning to school.
Myers, Starobin, Chen, Baul, and Kollasch (2015)	Aimed to understand the influence of community college students' engagement on their intentions to transfer and pursue a STEM degree. They investigated two questions. (1) How can student engagement be measured? (2) To what extent do student engagement and demographics influence students' decisions to major in a STEM degree?	Data was collected from a STEM Student Success Literacy Survey that was administered to all 15 community colleges in Iowa and analyzed using descriptive and inferential statistics including exploratory factor analysis. Logistic regression was used for predictive modeling.	• Nine variables were used for their predictive ability for students' intent to pursue STEM degree – level of science, level of math, native language, age, gender, concern for finances, number of hours worked weekly, highest degree desired, and intention to transfer. • This study revealed no significant impact of student engagement on STEM aspirations.
Lopez and Jones (2017)	Examined the experiences of students that transferred from a community college to pursue a STEM degree at a Midwestern Research University, while also looking at the academic and social factors that influenced their success.	Variables were examined with the Laanan Transfer Students' Questionnaire (L-TSQ) which captures background information, community college experiences, and university experiences. Descriptive statistics were used to provide a student profile, while regression analysis was used to predict student GPA and adjustment based on the variables captured in the L-TSQ.	• Prediction of student adjustment was indicated by father's highest level of education, community college experiences with faculty, university experiences with faculty, and negative perception as a transfer student. • GPA prediction was based on father's highest level of education, associate degree completion, transfer GPA, and total transfer credits.

Table 2. (*continued from previous page*)

post-secondary education at a community college. Unfortunately, FGCS often struggle with the same barriers as women and URM, including factors such as underprepared, work demands, lack of support, and high attrition rates (Dika & D'Amico, 2016). When investigating S&E graduates, it was found that older graduates were more likely to attend community college than younger students (Tsapogas, 2004). Overall, the community college student has a very different demographic than traditional college students and cannot be viewed through the same research lens (Costello, 2012; Horn & Nevill, 2006).

Given the increasing number of students attending community colleges, including racial minorities, it is important to investigate retention at these institutions (Starobin & Laanan, 2010). Tinto (1987) recognized that withdrawal rates were lowest among two year institutions and connected this low withdrawal rate to some of the various factors. The primary reasons for community college withdrawal rates being higher seems to be related to the lack of preparedness of students as well as students coming from a lower socioeconomic background (Cohen et al., 2014; Tinto, 1987). Hagedorn and DuBray (2010) studied a large cohort of community college students in California and found only 12.6% of the STEM-focused transfer-hopeful students were enrolled in a transfer level math, with the rest of the hopefuls being in lower remedial courses. The research also found success in math classes was significantly related to demographic data such as gender and race. The factors that impact student success for traditional university students might not be the same factors that community college students face, especially when considering women and minorities (Hagedorn & DuBray, 2010). Therefore, it is certainly worthy of investigation. Higher education students are no longer one-size-fits all, and the predictive analytical tools cannot be universal either.

Another area of concern is the lack of attention predictive models give to institutional factors. Given the importance of institutional factors identified by Tinto (1987), Astin (1993), and Seymour and Hewitt (1997), it is surprising that more recent STEM studies have continued to largely focus on student attributes. Some student retention studies investigate the importance of institutional factors, but are not usually concentrated on STEM education. For instance, Webster et al. (2011) investigated institutional factors in predicting student retention and found that tuition, student-teacher ratio, and the amount of aid received all influence a student in their decision to persist. This study also found a positive relationship between faculty salaries and retention, which reinforces the idea that more selective institutions have higher retention rates. Seymour and Hewitt (1997) repeatedly heard from students that the STEM educational system was designed to weed out minorities and lower socioeconomic students. The institutional diversity among community colleges needs to be investigated further to ascertain which institutional model is most successful for increasing STEM majors and diversity.

In the review of literature, there were some models aimed at identifying attrition causes and developing predictive models based on the data. There was a dearth of studies specifically investigating STEM students though, as Table 2 highlights.

When looking at predictive models there are some alarming limitations, one of which is the lack of a large breadth of research on the retention causation factors at the community college level. Community colleges are educating more students than ever and a majority of those are transfer students (Hagedorn & DuBray, 2010). It is reported that approximately half of all students receiving a STEM bachelor degree attended a community college for courses as undergraduates, but little research is being done to determine the factors contributing to the extremely low retention rates at two year colleges for STEM majors. There are many predictive models for student success and retention that provide strong evidence of causation factors, but few effectively transfer to the community college model.

V. Future Opportunities to Align Predictive Models with Community Colleges

There is a large effort to increase STEM retention. Many colleges and universities have invested in programs to support STEM students more effectively. The National Science Foundation (NSF) has developed grant opportunities to fill many of these deficiencies. Learning communities and faculty engagement have been shown to increase persistence by allowing students to make those important connections (Tinto, 1998, 2015). Louisiana State University developed a framework to show that student retention is clearly impacted by mentoring and undergraduate research. Their program specifically targeted academic underperformers and minorities (Wilson et al., 2011). NSF's S-STEM grant has provided institutions the ability to award scholarships and impact recruitment and retention. One institution had remarkable results by focusing on two factors: financial assistance and giving students a sense of belonging to STEM using various engagement strategies (Jen-Mei, Chuhee, Stevens, & Buonora, 2016). In addition, there are several collaborative efforts between community colleges and their transfer institutions that have promise. The Committee on Enhancing the Community College Pathways to Engineering found that the community college transfer function is critical to increasing and diversifying the workforce by enhancing the pathways through stronger articulation agreements and 2 + 2 plans (2005). NSF's Science Talent Expansion program works across the educational landscape to increase participation using pathways and transitional frameworks. It seems there are efforts to increase retention; however, community college students still do not align with many of the predictive tools being produced currently.

The development of predictive models and data analytics is gaining favor with educational researchers. There are multiple attempts to discern the best model for STEM students, but the models do not align with the community college student population. Most of the models include high school performance data, which might not be the best indicator for non-traditional students. The models that have been developed could be used with community college data to determine the efficacy. Additionally, there could be new models developed using a variety of techniques beyond the traditional regression analysis. When reviewing the research, engineering educational researchers have been the most creative in generating predictive models. The limitations of their models are related to the use of data from traditional universities. Future work should include validation tests using community college student data, as well as attempts to develop models based on the data from community colleges. Through a more holistic approach to predictive models, the problem surrounding STEM attrition could have clarity.

References

Alkhasawneh, R., & Hargraves, R. H. (2014). Developing a Hybrid Model to Predict Student First Year Retention in STEM Disciplines Using Machine Learning Techniques. *Journal of STEM Education: Innovations and Research, 15*(3), 35-42.

Astin, A. W. (1993). *What matters in college?: four critical years revisited* (Vol. 1st). San Francisco: Jossey-Bass.

Bernold, L. E., Spurlin, J. E., & Anson, C. M. (2007). Understanding Our Students: A Longitudinal-Study of Success and Failure in Engineering With Implications for Increased Retention. *Journal of Engineering Education, 96*(3), 263-274. doi:10.1002/j.2168-9830.1997.tb00259.x.

Burtner, J. (2005). The Use of Discriminant Analysis to Investigate the Influence of Non-Cognitive Factors on Engineering School Persistence. *Journal of Engineering Education, 94*(3), 335-338. doi:10.1002/j.2168-9830.2005.tb00858.x

Chang, M. J., Eagan, M. K., Lin, M. H., & Hurtado, S. (2011). Considering the Impact of Racial Stigmas and Science Identity: Persistence Among Biomedical and Behavioral Science Aspirants. *The Journal of Higher Education, 82*(5), 564-596.

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

Cohen. (2005). UCLA community college review: Why practitioners and researchers ignore each other (even when they are the same person). *Community College Review, 33*(1), 51-62.

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

Committee on Prospering in the Global Economy of the 21st, C., National Academy of Engineering Institute of, M., & National Academy of, S. (2007). *Rising Above The Gathering Storm: Energizing and Employing America for a Brighter Economic Future:* National Academies Press.

Committee on Underrepresented, G., the Expansion of the, S., Engineering Workforce, P., Committee on Underrepresented, G., the Expansion of the, S., & Engineering Workforce, P. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads.* Washington, D.C: National Academies Press.

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

Council, N. R., Division on, E., Physical, S., Board on Mathematical, S., Their, A., & National Research Council . Committee on the Mathematical Sciences, i. (2013). *The mathematical sciences in 2025.* Washington, D.C: National Academies Press.

DeBerard, M. S., Spelmans, G. I., & Julka, D. L. (2004). Predictors of Academic Achievement and Retention Among College Freshmen: A Longitudinal Study. *College Student Journal, 38*(1), 66-80.

Deegan, W. L., Tillery, Dale. (1985). *Renewing American community college: priorities and strategies for effective leadership.* San Francisco: Jossey-Bass.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems, 49*(4), 498-506. doi:10.1016/j.dss.2010.06.003

Dika, S. L., & D'Amico, M. M. (2016). Early experiences and integration in the persistence of first-generation college students in STEM and non-STEM majors. *Journal of Research in Science Teaching, 53*(3), 368-383. doi:10.1002/tea.21301

Djulovic, A., & Li, D. (2013). Towards Freshman Retention Prediction: A Comparative Study. *International Journal of Information and Education Technology, 3*(5), 494. doi:10.7763/IJIET.2013.V3.324

Drew, D. E. (2011). *STEM the tide: reforming science, technology, engineering, and math education in America.* Baltimore: Johns Hopkins University Press.

Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. (2012). President's Council of Advisors on Science and Technology.

Eris, O., Chachra, D., Chen, H. L., Sheppard, S., Ludlow, L., Rosca, C., . . . Toye, G. (2010). Outcomes of a Longitudinal Administration of the Persistence in Engineering Survey. *Journal of Engineering Education, 99*(4), 371-395. doi:10.1002/j.2168-9830.2010.tb01069.x

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis.* Oxford;New York;: Oxford University Press.

French, B. F., Immekus, J. C., & Oakes, W. C. (2005). An Examination of Indicators of Engineering Students' Success and Persistence. *Journal of Engineering Education, 94*(4), 419-425. doi:10.1002/j.2168-9830.2005.tb00869.x

Hagedorn, L. S., & DuBray, D. (2010). Math and science success and nonsuccess: Journeys within the community college. *Journal of Women and Minorities in Science and Engineering, 16*(1).

Hagedorn, L. S., & Purnamasari, A. V. (2012). A Realistic Look at STEM and the Role of Community Colleges. *Community College Review, 40*(2), 145-164.

Hall, C. W., Kauffmann, P. J., Wuensch, K. L., Swart, W. E., DeUrquidi, K. A., Griffin, O. H., & Duncan, C. S. (2015). Aptitude and Personality Traits in Retention of Engineering Students. *Journal of Engineering Education, 104*(2), 167-188. doi:10.1002/jee.20072

Herzog, S. (2006). Estimating Student Retention and Degree-Completion Time: Decision Trees and Neural Networks Vis-a-Vis Regression. *New Directions for Institutional Research, 2006*(131), 17-33. doi:10.1002/ir.185

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

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

Jackson, D. L., & Laanan, F. S. (2011). The role of community colleges in educating women in science and engineering. *New Directions for Institutional Research, 2011*(152), 39-49. doi:10.1002/ir.407

Jen-Mei, C., Chuhee, K., Stevens, L., & Buonora, P. (2016). Strategies to Recruit and Retain Students in Physical Sciences and Mathematics on a Diverse College Campus. *Journal of College Science Teaching, 45*(3), 14-22.

Li, Q., McCoach, D. B., Swaminathan, H., & Tang, J. (2008). Development of an Instrument to Measure Perspectives of Engineering Education Among College Students. *Journal of Engineering Education, 97*(1), 47-56. doi:10.1002/j.2168-9830.2008.tb00953.x

Li, Q., Swaminathan, H., & Tang, J. (2009). Development of a Classification System for Engineering Student Characteristics Affecting College Enrollment and Retention. *Journal of Engineering Education, 98*(4), 361-376. doi:10.1002/j.2168-9830.2009.tb01033.x

Luan, J. (2002). Data Mining and Its Applications in Higher Education. *New Directions for Institutional Research, 2002*(113), 17-36. doi:10.1002/ir.35

Marra, R. M., Rodgers, K. A., Shen, D., & Bogue, B. (2012). Leaving Engineering: A Multi-Year Single Institution Study. *Journal of Engineering Education, 101*(1), 6.

Montgomery, D. C., Vining, G. G., & Peck, E. A. (2012). *Introduction to linear regression analysis* (Vol. 821). Hoboken, NJ: Wiley.

Morganson, V. J., Major, D. A., Streets, V. N., Litano, M. L., & Myers, D. P. (2015). Using Embeddedness Theory to Understand and Promote Persistence in STEM Majors. *Career Development Quarterly, 63*(4), 348-362. doi:10.1002/cdq.12033

National Science Board. (2016). *Science and Engineering Indicators 2016*. Arlington, VA: National Science Foundation.

Nicholls, G. M., Wolfe, H., Besterfield-Sacre, M., Shuman, L. J., & Larpkiattaworn, S. (2007). A Method for Identifying Variables for Predicting STEM Enrollment. *Journal of Engineering Education, 96*(1), 33-44. doi:10.1002/j.2168-9830.2007.tb00913.x

Olsen, S. (2014). *The importance of engineering talent to the prosperity and security of the nation: summary of a forum.* Washington, District of Columbia: National Academies Press.

Osborne, J. W. B., Erin S. (2016). *Exploratory Factor Analysis with SAS* (Vol. 1): SAS Institute.

Packard, B. W.-L., Gagnon, J. L., LaBelle, O., Jeffers, K., & Lynn, E. (2011). Women's Experiences in the STEM Community College Transfer Pathway. *Journal of Women and Minorities in Science and Engineering, 17*(2), 129-147. doi:10.1615/JWomenMinorScienEng.2011002470

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: why undergraduates leave the sciences*. Boulder, Colo: Westview Press.

Sislin, J., Mattis, M. C., & National Research Council . Committee on Enhancing the Community College Pathway to Engineering, C. (2005). *Enhancing the community college pathway to engineering careers.* Washington, D.C: National Academies Press.

Snyder, T. D., Dillow, S. A., & Staff, I. o. E. S. (2012). *Digest of Education Statistics 2011:* U.S. Government Printing Office.

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

Starobin, S., & Laanan, F. (2008). Broadening female participation in science, technology, engineering, and mathematics: Experiences at community colleges. *New Directions for Community Colleges, 2008*(142), 37-46. doi:10.1002/cc.323

Swanger, D. (2013). *Community colleges: partners in community development : approaches to developing our regions:* Society for College and University Planning

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

Tinto, V. (1987). *Leaving College: Rethinking the Causes and Cures of Student Attrition*: The University of Chicago.

Tinto, V. (1998). Colleges as Communities: Taking Research on Student Persistence Seriously. *Review of Higher Education, 21*(2), 167.

Tinto, V. (2015). Access without support is not opportunity. *Community College Week,* p. S36. Retrieved from http://mst.summon.serialssolutions.coms

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*(3), 207-222.

Tsapogas, J. (2004). The role of community colleges in the education of recent science and engineering graduates. *InfoBrief NSF*, 04-315.

Veenstra, C. P., Dey, E. L., & Herrin, G. D. (2008). Is Modeling of Freshman Engineering Success Different from Modeling of Non-Engineering Success? *Journal of Engineering Education, 97*(4), 467-479.

Vilorio, D. (2014). STEM 101: intro to tomorrow's jobs. *Occupational Outlook Quarterly, 58*(1), 2.

Wilson, Z. S., Holmes, L., deGravelles, K., Sylvain, M. R., Batiste, L., Johnson, M., . . . Warner, I. M. (2011). Hierarchical Mentoring: A Transformative Strategy for Improving Diversity and Retention in Undergraduate STEM Disciplines. *Journal of Science Education and Technology, 21*(1), 148-156. doi:10.1007/s10956-011-9292-5

Xue, Y., & Larson, R. C. (2015). STEM crisis or STEM surplus? Yes and yes. *Monthly Labor Review*, 1B.

Young, J. W. (1997). Community Economic Development Through Community Colleges. *New Directions for Higher Education, 1997*(97), 74-83. doi:10.1002/he.9707

Jennifer Snyder is Dean of Science for East Campus at Valencia College in Orlando, Florida. Prior to that position, she was Department Chair of Physical Science and Chemistry Faculty at Ozarks Technical Community College in Springfield, Missouri. She received her Bachelor of Science in Chemistry from Southwest Missouri State University and Master of Science in Chemistry from Missouri State University. She is a PhD Candidate in the Engineering Management and Systems Engineering Department at Missouri University of Science and Technology.

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