5-2016

Pre-college Characteristics and Online Homework Learning: Factors Associated with First Year Engineering Students' Academic Success

Wenshu Li

University of Tennessee - Knoxville, wli23@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Wenshu Li entitled "Pre-college Characteristics and Online Homework Learning: Factors Associated with First Year Engineering Students' Academic Success." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Educational Psychology and Research.

Gary Skolits, Major Professor

We have read this dissertation and recommend its acceptance:

Jennifer Morrow, Richard Bennett, Wenjun Zhou

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Pre-college Characteristics and Online Homework Learning: Factors Associated with First Year Engineering Students' Academic Success

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Wenshu Li

May 2016

Dedication

To my mom, cousin, aunt, and all other close relatives, whose love, encouragement, support, and inspiration made me enjoy learning and able to pursue my career goals.

To my husband, who always supports me with his unconditional love and companionship through my way to achieve the success.

Acknowledgement

I would like to express my deepest gratitude to my chair and advisor, Dr. Gary J. Skolits. Thank you for your excellent guidance, caring, patience, and encouragement through my entire graduate life and dissertation process.

To my committee members, Dr. Jennifer Ann Morrow, Dr. Richard Bennett, and Dr. Wenjun Zhou, millions of thanks for your expertise, constructive comments, support, and time. I could not come to here without your endless support and guidance.

I would like to express my appreciation for Dr. Richard Bennett, Dr. Rachel McCord, Mr. William Schleter, and all Teaching Assistants of EF 151 course in my process of data collection. I'm well aware of the difficulty of meeting IRB requirements on data collection especially when I have to work full time out of town and having data collection on campus. Thank you very much to spend your valuable time helping me distributing the consent form, sending survey recruitment emails, and preparing the online activity logs datasets.

I must thank my colleagues and friends in the Evaluation, Statistics, and Measurement program. Thank you for supporting me through the past 4 years with your individual skills and expertise.

Last but not the least, I want to extend my gratitude to my family and friends, who are always there for me. You will likely never fully know how much you each mean to me.

Abstract

The purpose of the study was to develop a working model to predict at risk students in an Introduction to Engineering course. The model considers both students' pre-college characteristics, psychological traits, and online homework learning behavior. The study assisted the course instructor in the creation of an early warning system and the development of targeted interventions for students at risk. A reliable and valid instrument to measure engineering students' pre-college characteristics was initially developed. The study also applied data mining to analyze the student online homework logs in order to observe engineering students' homework learning process. A decision tree model containing all of the pre-college characteristics and online homework learning features was also developed, and it identified four key factors related to students' risk to fail the first module exam: Correctness, Preparedness, Self-efficacy, and percentage of homework attempts after deadline (Plate). The results of the decision tree model helped identify students-at-risk at early stage of the course. Students at risk were grouped into multiple groups. The author also proposed customized interventions to help students in different at risk groups. The findings of the study helped engineering students and educators to build up a comprehensive student profile to better understand students' academic status and learning needs in the course. Thus this study suggests ways for both the engineering educators and students to improve the learning process in a more efficient manner.

Table of Contents

Chapter 1 Introduction 1
Statement of the Problem 2
Purpose of the Study 6
Hypothesis 6
Research Questions 6
Definition of terms 7
Limitation 9
Organization of the Study 10
Chapter 2 Literature Review 11
Pre-college Characteristics Related to First Year Engineering Students' Success 11
First Year Engineering Students' Learning Process 18
First Year Engineering Success/Retention Model 21
Psychometric Properties of Instruments 25
Data Mining within A Course Management System 27
Conclusion 27
Chapter 3 Method 33
Purpose of the Study 33
Hypothesis 33
Research Questions 33
Research Design 34
Population and Sampling 34
Participants 34
Data collection process 35
Survey Instrument 40
Building Decision Tree Model 42
Chapter 4 Results 45
Data Cleaning 46
Hypothesis: Survey Validation 47
Research Question 1: How actively are first year engineering students engaged in their online homework study? 55
Research Question 2: Are there any group differences (gender and first generation) on students' pre-college characteristics, psychological traits, and online homework activities? 55

Research Question 3: What factors are associated with students' risk to fail a course? 56
Research Question 4: What are the characteristics of students at risk? 58
Chapter 5 Discussion 62
Validity and Reliability of the Engineering Students' Pre-college Characteristics Survey 62
Research Question 1: How actively are first year engineering students engaged in their online homework study? 64
Research Question 2: Are there any group differences (gender and first generation) on students' pre-college characteristics, psychological traits, and online homework activities? 66
Research Question 3: What factors are associated with students' risk to fail a course? 67
Research Question 4: What are the characteristics of students at risk? 70
Practical Implication 71
Limitations of the Present Study 74
Future Research 75
Conclusion 76
List of References 77
Appendices 89
Appendix A 90
Appendix B 91
Appendix C 101
Appendix D 103
Appendix E 106
Vita 107

List of Tables

Table 1. Psychometrics Properties of Survey Instrument (Furr & Bacharach, 2014)……………………………………………..…………………………28

Table 2. Participants' Demographics…………………..……………………………39

Table 3. Fitting Indices of Hypothesized Model…………………...………………49

Table 4. Item Loadings of The hypothesized model……………………………….51

Table 5. Fitting Indices of Modified Model………………..……………………….52

Table 6. Standardized Item Loadings of the New Model…………………………..53

Table 7. Group Comparison on Pre-college Characteristics and Homework Learning Activity Features

……………………………………………………………………………………57

Table 8. Summary of Leaf Nodes……………………………………………….….61

List of Figures

Figure 1. Engineering Retention Model by Veenstra, Dey, and Herrin (2009)…….24

Figure 2. The Process of Data Mining ……………………..…………………....38

Figure 3. Example of A Decision Tree……………………………………………..44

Figure 4. Hypothesized Instrument Structure………..……………………………50

Figure 5. Modified Survey Instrument Structure…………………………...……54

Figure 6. Decision Tree Model………………………………..………………60

Chapter 1

Introduction

This chapter provides an overview of the study. It briefly describes the significance, the framework of the study, and the manner in which the author conducted the study.

The report, The Engineer of 2020 (Clough, 2004), introduced a possible vision of Engineering in 2020. This report introduced the expectation that engineers need to be equipped with the following knowledge and skills sets: strong analytic skills, practical ingenuity, creativity, communication, principles of business and management, leadership, sense of professionalism, dynamism, agility, resilience, flexibility, and become a lifelong learner.

With these high expectations for engineers in 2020, engineering education is currently facing serious challenges including an inability to address the market demand for engineering jobs (Weiss, 2009), lack of diversity in engineering workforce (Weiss, 2009), a decrease in the number of engineering graduates (Ohland, Sheppard, Lichtenstein, Eris, Chachra, & Layton, 2008), and a decrease in engineering graduates' persistence in engineering careers (Jones, Paretti, Hein, & Knott, 2010). According to the U.S. National Science Foundation (NSF), enrollment in undergraduate engineering programs in the United States dropped while the market demand for engineering students increased (National Science Foundation, 2004). A recent study using data from the National Student Clearinghouse found that only 24.5% of the White students and 32.4% of Asian American students who entered college majoring in STEM degree areas actually achieved a STEM degree in four years. By contrast, only 15.9% of Latino students, 13.2% of Black students, and 14.0% of Native American received STEM degrees in four years (Eagan, Hurtado, & Chang, 2010). Even if the students who majored in Engineering obtained their Engineering degree, it would not necessarily ensure that these graduates would subsequently pursue engineering careers (Jones, Paretti, Hein, & Knott, 2010).

To address these challenges, engineering educators and researchers have devoted themselves to improving engineering education and cultivating qualified engineers more effectively. To be specific, more engineering professors have started to investigate literature about engineering education and some are attending engineering education related conferences and teaching workshops (Rugarcia, Felder, Woods, & Stice, 2000). While engineering researchers and professors focus on college education and the development of a variety of approaches to engage college students, they are also concerned about pre-college engineering education (Phase, 2005). Engineer educators have noted that the current K-12 system does not adequately prepare students for entering and succeeding in engineering programs (Phase, 2005).

In summary, both K-12 and college education are burdened by the responsibilities to prepare future engineers. It is important to understand college engineering students' success taking consideration of both their pre-college background and college effort.

Statement of the Problem

There are several first year success and/or retention models focusing on college student success. These would include Tinto's interactionalist theory model (Tinto, 1993) and Astin's theory of involvement model (Astin, 1984). However, there are relatively few models specifically developed for first year engineering students. Engineering students are different from their cohorts in social science, humanities, arts, and even other STEM disciplines in many aspects such as social engagement and their SAT/ACT math scores (Veenstra, Dey, & Herrin, 2008). As a result, the first year success models focused on all college students do not necessarily apply directly to engineering student. Therefore, it is necessary to develop a new model or expand the existing first year success model to address first year engineering students' success.

Veenstra, Dey, and Herrin (2008) developed a first year engineering student success model through factor analysis and linear regression. This model was based on Tinto's interactionalist theory (Tinto, 1993). The model effectively addressed engineering students' pre-college characteristics and the relationship with first year academic success. Unfortunately, the study failed to assess the psychometrical properties of the instruments as well as relevant aspects of students' integration processes such as learning and social integration at college. This essentially created a "black box" model that only takes into account students pre-college characteristics and end of first year academic outcome and neglecting the first year learning process.

With the explosion of information technology innovations and the increasing use of the Internet, educational institutes around the world are changing their methods of delivering knowledge (Tella & Tella, 2011). Universities and colleges are transiting from traditional face-to-face classroom teaching to online teaching by offering online courses and promoting e-learning through the use of online Course Management Systems (CMSs) such as Moodle (Moodle, 2007) and Blackboard (BlackBoard, 2007). The use of CMSs enable the educators and researchers to observe students' learning process by tracking students' activities logs (Minaei-Bidgoli, Kashy, Kortemeyer, & Punch, 2003; Romero, Ventura, & García, 2008).

In order to better engage first year engineering students through effective teaching approaches and interventions at the University of Tennessee, Knoxville, a program, entitled *Engage,* was developed in 1999. The program consists of two 4-hour courses (Physics for Engineers I and Physics for Engineers II) and a 1-hour computer course (Computer Methods in Engineering Problem Solving). Students enrolled in the university intending to pursue a major in engineering are required to take this program. The Physics courses combined the traditional lecture with an inquiry-based and project-based teaching approach. The class format is large (approximately 500 students) 50-minute lectures three days a week and smaller (24-36 students) 75-minute recitations two days a week. The lectures are team-taught and use a personal response system to increase course engagement. The recitations are led by trained graduate assistants and consist of collaborative problem solving, hands-on activities, demonstrations, and team projects.

The courses use a customized web-based homework system (Schleter & Bennett, 2006). This system provides individualized online homework (Goulet, 2010). Each student has different parameters so that they will have different homework problem sets, which means students will work on same homework problems but the numbers in the questions will be different. In order to increase students' engagement in homework, a bonus system (Schilling, 2010) was implemented in the homework system in 2010, enabling students to receive a 10% bonus for homework problems completed at least 24 hours or more in advance of the due date. The bonus system has resulted in over half of the homework being completed within the bonus time (Bennett, Schleter, & Raman, 2012). The bonus system encourages self-taught student learning, discourages procrastination, improves preparation for lecture and recitation, and reinforces learning (Bennett, Li, Olsen, & Schleter, 2013).

This online homework system keeps track of students' activities features, including when students log into the system, how many times they attempt to solve certain questions, and whether or not they answer the question correctly, which could help researchers to observe students' homework learning process.

In summary, the existing models related to first year engineering students' academic success were limited to exploring students' pre-college characteristics using survey instruments. Even though some of the models suggested incorporating students' college integration process (leaning and social), few studies collected and analyzed data on both pre-college characteristics and college integration. It is important to develop a working model to predict first year engineering students' academic success considering both pre-college characteristics and college integration process.

The increasing use of course management system enables researchers to collect observation data on students' learning activities. Instead of having students self-report their engagement on academic studies (i.e., how many hours student spent on learning activities), researchers can apply data mining technique to analyze the observational data recorded in the course management logs and describe students' activities. Specifically in the *Engage* program online homework system, researchers can observe students activities in variety of aspects, for example, how much time student spent and how many attempts students have on each question, problem set, homework set, and when students start working on their homework, during what time of a day. By applying data mining techniques on the vast volume of real-time homework activity data, researchers can get a more detailed and vivid picture of students' homework learning activities.

Purpose of the Study

In the present study, the author investigated students' pre-college characteristics such as exposures to engineering education and professionals, and psychological traits such as self-confidence in math and science skills, motivation of studying in Engineering. The author also applied data mining techniques to describe students' engineering homework learning activities. Framed by Tinto's interactionalist theory model (1993) and Veenstra et al.'s first year engineering retention model (2009), the author aimed to create a working model to predict students at risk in an Introduction to Engineering course by considering both students' pre-college characteristics, psychological traits, and online homework learning behavior. In this way, this study can help the course instructor to create an early warning system and develop targeted interventions for students at risk.

Hypothesis

The Engineering Students' Pre-college Characteristics survey demonstrates an acceptable internal consistency reliability (Cronbach's alpha >.60) and validity (construct).

Research Questions

1. How actively are first year engineering students engaged in their online homework study?
2. Are there any group differences (gender and first generation) on students' pre-college characteristics, psychological traits, and online homework activities?
3. What factors are associated with students' risk to fail a course?

4. What are the characteristics of students at risk?

Definition of terms

Psychometric Properties.

Reliability: a test/survey gets consistent result under consistent conditions over time.

Test-retest reliability: Involves administering the same test twice to the same group after a certain time interval has elapsed. Reliability is calculated by the correlation between two sets of data.

Equivalent form reliability: Two different but equivalent (alternate or parallel) forms of an instrument are administered to the same group during the same time period.

Internal-consistency methods: include calculating cronbach's alpha and split half procedure.

Validity: the extent to which the instrument measures what it intends to measure

Translation validity: includes face validity and content validity. Face validity is the weakest way to measure validity. To judge whether or not the test has face validity, just read the items and subjectively judge whether or not the items measure the construct. Content validity addresses the match between test questions and the content they are intended to measure.

Criteria validity: includes two types of validity: predictive validity and concurrent validity. Predictive validity compares the instrument scores and criterion scores at a later time while concurrent validity compare instrument data and criterion data at the same time.

Construct validity: how well the instrument measures the theoretical constructs it intends to measure.

Students' Pre-college characteristics.

Engineering study skills: students' learning strategies that can help them understand and use the course content.

Motivation: this variable is related to why students choose engineering as their college major.

Students' family/social background: this section of variables relates to students' family income and parental education level.

Students' self-efficacy in Math and Science skills: students' self-beliefs on their competences to solve math and science related problems.

Students' knowledge about engineering professions: this variable measures how much students know about engineering careers and engineering education.

Students' learning process.

Students' first year learning process is a very broad concept and on-going process. In this study, we are focused on first year engineering students' online homework learning process in an introductory engineering course.

Pearly, Pregular, Plate: percentage of homework assignments accomplished during bonus period, regular period (i.e., the day when the homework is due), and after deadline (i.e., late).

PansweredQ: percentage of homework problems the student has answered. This variable is calculated by dividing the number of problems a student has had attempts on by the total number of available problems in the homework assignment.

Mattempts: the average number of attempts per problem. This variable is calculated by dividing the total number of a student's attempts by the total number of problems.

Correctness: percentage of problems solved correctly. This variable is calculated as the total number of a student's correct answers divided by the total number of problems.

LearningSpeed: the average number of attempts per correct answer. The LearningSpeed variable may be interpreted as the number of attempts a student needs on average to find the correct answer. It is calculated by dividing the total number of attempts of a student by the total of correct answers.

Preparation: the percentage of problems a student solves correctly on the first attempt. For each of the homework problem, if a student can answer it correctly without hint or incorrect attempts, we believe the student is well prepared for the homework. A student may be prepared by in-class learning through lectures, or collaborative-learning through group study.

Students' academic success.

In the present study, students' academic success will be measured by students' exam 1 scores in an introductory engineering course, EF151.

Limitation

The major limitation of the study is the sample size and impact of fulfilling the Institutional Review Board (IRB) requirements on participants' recruitment process. One of the threats to the power of statistical testing is the small sample size (Shadish, Cook, & Campbell, 2002). For example, one of the planned statistic analyses in the current study was confirmatory factor analysis, which required the sample size over 300. With the population of 585 first year engineering students enrolled in EF151, this required the response rate to online survey over 51.3%.

One of the other data collection methods in the study was data mining of students' online homework activity logs. As required by IRB, the research has to obtain each student's written permission/consent in order to request the activity logs from the course instructor and this process must be separate from the online survey consent process.

In order to maximize students' participation rate in the study and collect as much data as possible, the researcher worked together with the course instructor to provide incentives (extra credit) for students' online survey participation. A separate written consent form was distributed by teaching assistant to students and ask for students' permission to release their online homework activity logs. I also adopted a statistical modeling approach, decision tree modeling, which does not have strict requirement on sample size to proceed with the analysis and model building process.

Organization of the Study

This study is organized into five chapters. The current chapter, chapter 1, introduced the study background, stated the problem and purpose of the study, provided research questions and definitions of concepts, and reviews the organization of the chapters. Chapter 2 provides a review of the literature, including the framework and gaps in the literature, and the significance of the study as why it is important to the field. Chapter 3 describes the research methods including information on research design, procedure of data collection, and participants. Chapter 4 depicts the results and findings of the study. Chapter 5 concludes the research with an overall discussion of the results, practical implication, limitation, and future studies.

Chapter 2

Literature Review

This chapter provides a synthesis and critique of the theoretical and empirical literature on the factors that contribute to first year engineering students' academic success. It has been organized into 5 sections. The chapter began with a synopsis of engineering students' pre-college characteristics. Then, the chapter described related research about students' first year learning process. Thirdly, theoretical models on engineering students' success were synthesized. Then the chapter reviewed the psychometrics properties about survey instrument. Lastly, the chapter discussed data mining in educational studies.

Pre-college Characteristics Related to First Year Engineering Students' Success

Engineering study skills.

Engineering study skills have been identified as important pre-college characteristics of engineering students (Bernold, 2007; Haase, Chen, Sheppard, Kolmos, & Mejlgaard, 2013; Veenstra et al., 2009). Some of the important skills identified by Bernold were stress management, test taking, note taking, reading to learn, time management, and meta-cognition awareness. In the study to investigate first year engineering students' baseline of engineering study skills, 1,020 first year engineering students were asked to respond to the Learning and Study Strategies Inventory (LASSI). Bernold (2007) found that time management was the skill that most engineering students were lacking. The other two areas of weakness were knowledge of study methods/aids and comprehension monitoring techniques. Veenstra et al. (2009) believed that first year engineering students with good study habits such as independent learning will have better academic achievement in their first year engineering study.

In a large-scale and cross-national study, researchers compared US and Denmark students' perceived importance on both interpersonal and professional skills, and math/science skills (Haase, Chen, Sheppard, Kolmos, & Mejlgaard, 2013). The researcher divided engineering study skills into two major categories: interpersonal and professional skills (IPP) and math/science skills (MS), and they established a quadrant model based on these two aspect of skills. Students were then divided into four groups: double focus group who consider both IPP and MS skills were important, MS focus group who overweighed MS skills than IPP skills, IPP focus who believed that IPP skills were more important, and Not impressed group who thought neither of the skills were important. US and Denmark engineering students were compared in each dimension of the quadrant model. The author also identified the characteristics of each group regarding their demographic information and motivation.

Student' motivation.

Motivation to succeed in engineering is also considered as an important pre-college characteristic of engineering students (Lotkowski, Robbins, & Noeth, 2004). There are different types of motivation theories and researchers conducted studies investigating the relationship between different types of motivation constructs and engineering academic outcomes. For example, the expectance-value theory predicts that student performance is directly influenced by both students' expectancies for success and ability-related beliefs (Eccles et al., 1983). French, Immekus, and Oakes (2005) reported strong correlations between retention and motivation in their study. Motivation, measured by the Academic Intrinsic Motivation Scale, was a significant predictor of students' enrollment in the university and engineering. Their study found that a higher level of motivation was significantly related to continuing in the major. As defined by the expectance-value theory, the achievement task value consisted of at least three parts: attainment value, intrinsic value or interest, and extrinsic utility or value (Eccles & Wigfield, 1995). Attainment value refers to the importance of achievement in engineering in terms of individual core values. Intrinsic value is students' interest in engineering activities and extrinsic utility is the usefulness of engineering to achieve personal goals.

Ahmad et al.'s study (2012) considered students' interest in engineering study as one of the most important pre-program preparedness characteristics. This is in contrast to French et al.'s study (2005) that utilized a published questionnaire to investigate students' motivation. They developed seven 5-point Likert scale items to investigate both students' internal, external and attainment motivation of studying in the engineering program. The sample items included "Profound interest in electrical/ mechatronics engineering" and "Career prospects of the programs". The 7 items motivation survey had a favorable internal reliability with Cronbach's alpha equal to 0.845. The result suggested that students had high interest in the program of study, with mean score at every item over 3.3 on a 5-point scale. This result indicated that first year students in the engineering program have prepared themselves with the needed intention to study in the program.

Also based on the expectancy-value theory, Li, McCoach, Swaminathan, and Tang (2008) developed an instrument to investigate engineering and non-engineering students' motivation. The instrument has three subscales related to each component of expectancy-value theory: intrinsic value, utility value, and cost. The questionnaire had favorable content validity through expert review and it also had high internal reliability. Applying the instrument to investigate students' motivation, the authors found that although engineering is beneficial to the society, it is hard to achieve an engineering degree and that not much personal benefit was associated with pursuing an engineering degree.

While the survey developed by Li et al. (2008) was used to measure both engineering and non-engineering students' motivation, Sheppard et al. (2010) developed a motivation survey specifically for engineering students. In their survey, motivation was divided into five components: intrinsic, parental, social, financial, and mentor motivation through Principal Component Analysis (PCA). This survey reflected favorable reliability and validity characteristics. Haase et al. (2013) also applied this instrument in their large-scale study to measure US and Denmark engineering students' motivation for engineering study. The instrument also displayed good internal reliability in their study with Cronbach' alpha in most subscales over 0.75. The study found that Denmark and US engineering students are not different significantly in intrinsic, societal, and mentor motivation. However, US engineering students rated financial motivation more important while Denmark engineering students believed parental motivation was more important.

Students' self-efficacy on science and math skills.

According to Bandura's Social Cognition Theory (SCT), self-efficacy is defined as individual's self-belief in his/her competence to complete a task and achieve the desired goals (Bandura, 1986). A large collection of empirical studies has found the positive relationship between self-efficacy and students' academic outcome in college (Zajacova, Lynch, & Espenshade, 2005; Lotkowski, Robbins, & Noeth, 2004; Chemers, Hu, & Garcia, 2001). For example, Zajacova et al. developed a survey instrument to measure the level of academic self-efficacy and perceived stress associated with 27 college-related tasks. They also conducted factor analysis to validate the self-efficacy and stress constructs. Then they applied structural equation models to assess the importance of stress and self-efficacy in predicting academic performance outcomes such as first-year college GPA, the number of accumulated credits, and college retention after the first year. Their results indicated that self-efficacy rather than stress was a significant predictor of students' first year GPA. Similarly, Lotkowski et al.'s ACT policy reports (2004) posited the importance of self-efficacy on students' academic success. They found that academic self-confidence had strong relationship to college GPA.

Several empirical engineering studies also substantiated the importance of self-efficacy on students' academic success (Lent et al., 2008; Vogt, Hocevar, & Hagedorn, 2007). Based on Bandura's social cognition theory, Vogt et al. (2007) established a structural equation model to examine relationships among variables related to environment (discrimination and Academic integration), self (self-confidence and self-efficacy), behavior (help-seeking, peer learning, effort, and critical thinking), and academic outcome (GPA) using more than 700 engineering students as their sample. Self-efficacy, as one of the important self-related variables, was measured by the Motivated Strategies for Learning Questionnaire (MSLQ). The questionnaire included students' self-efficacy beliefs items in seven aspects: (1) master the skills in one's major; (2) understand both the most basic and (3) complex concepts; (4) understand the most difficult material in one's major; (5) will receive better than average grades in my major; (6) will do well in one's major; and (7) do an excellent job on assignments and projects within one's major. Their results indicated that students' self-efficacy have the strongest relationship with GPA among other variables.

Students' knowledge about engineering professions.

According to Hirsch, Kimmel, Rockland, and Bloom (2005, pS2F-21), "One of the many reasons more students are not choosing to study engineering in college and pursue careers in engineering is that they simply do not know what engineering is or what engineers do." In contrast to other careers such as doctors, counselors, and teachers, engineering is a major that is rarely talked about by teachers, media, and even parents. Hirsch et al. indicated that it might because parents, teachers, and school counselors don't know much about engineering majors and careers. Due to this lack of knowledge, they could not provide appropriate engineering related information to their children and students in regards of engineering as a college major and future career. In Hirsch et al.'s follow-up study (2006), they investigated high school students' knowledge about engineering by asking students to list five engineering careers. Less than 12% of the students could list 5 correct engineering careers. About 35% of the Students reported that their teacher never presented engineering principals as part of their classroom teaching.

Similarly, Knight and Cunningham (2004) posited that the public does not have a complete image of engineering professionals. Although surrounded by engineers in their daily life, students did not know exactly what engineers do. They believed that in order to increase students' knowledge about engineering careers, it is better to know students' image about engineering professions first. Hence they developed a Draw an Engineer Test (DAET) tool to investigate students' ideas about engineers and engineering. Their instruments asked students to spend 15 minutes writing and drawing what they thought about engineering. The researchers then recoded the written and picture responses into themes. Their results found that many students believed that engineers used tools to build buildings and fix car engines.

Students' knowledge about engineering professions is important because perceptions of careers play a key role in students' decision on whether or not they will enter into the careers. While academic preparation and motivation is essential for students to entering an engineering major in college, appropriate and detailed information about engineering and engineers should be disseminated to pre-college students.

Students' social/family background.

Students' background characteristics such as first generation status, education level of parents, financial needs, and family income are also important factors that influence first year engineering students' academic achievement and retention rate (Veenstra et al., 2009). It has been well documented that social economic status (SES) such as parental education level (Astin & Oseguera, 2005; Terenzini, Springer, Yaeger, Pascarella, & Nora, A., 1996) and financial needs (Wohlgemuth, et al., 2007) were related to students' academic achievement and retention rate.

Sheppard et al. (2010) examined engineering students' SES variables such as perceived family income and parental education level and their relationship with students' confidence in: 1) math and science skills, and 2) professional/interpersonal skills. They applied the Academic Pathways of People Learning Engineering Survey (APPLES) to examine students' SES and confidence level. They conducted two regression analyses using SES and other survey constructs (motivation, academic involvement, etc.) as predictors and students' confidence as dependent variable. Their results indicated that family income was important predictors of engineering students' confidence level of both math and science skills and professional/interpersonal skills.

First Year Engineering Students' Learning Process

While a tremendous amount of empirical studies exist addressing the influence of engineering students' pre-college traits on students' academic success, other empirical studies (Bernold, 2007; Liberatore, 2011) focused on improvements related to the engineering curriculum and teaching approaches to better engage first year engineering students in their academic study. For example, Bernold proposed that inquiry-based learning is "the pedagogical paradigm for 21st century" (Bernold, 2007). Compared to the traditional lectures that the professors spend their time filling students "empty brain", the inquiry-based learning requires students to manage their own learning. Students need to know the way of acquiring knowledge, developing personal strategies, recognizing their personal strength and weakness, and gaining new knowledge.

With the explosion of information technology innovations and the increasing use of the Internet, educational institutes around the world are changing the manner in which they deliver knowledge (Tella & Tella, 2011). Universities and colleges are transiting from traditional face-to-face classroom teaching to online teaching by offering online courses and promoting e-learning through the use of online Course Management Systems (CMSs). Educators are now making use of a variety of online CMSs, such as Moodle (Moodle, 2007) and Blackboard (BlackBoard, 2007), to distribute class materials, post course announcements, assign and grade assignments, and engage students with online discussion forums.

In particular, online homework within the CMSs is widely used to enhance student learning, especially in the science, technology, engineering and mathematics (STEM) areas, where quantitative computations are a major component of the homework assignments (Peng, 2009; Cheng, Thacker, Cardenas, & Crouch, 2004). Compared to traditional paper- and-pencil based homework, online homework has a number of advantages. Firstly, it helps reducing cost in collecting, grading and managing homework assignments. The automatic grading capacity of online homework makes it easy to manage a large number of students. Secondly, the interactive nature of online homework enables students to engage in and reflect on the learning process (Peng, 2009), which results in more effective learning. Finally, with randomly generated parameters for the same problem, computerized homework systems reduce chances of plagiarism. In one word, online homework has become a highly effective and efficient tool for teaching large classes of STEM students.

The University of Tennessee *Engage Program* incorporates inquiry-based and project-based teaching into the traditional lectures. The courses in the program use a customized web-based homework system (Schleter & Bennett, 2006). This system provides personalized homework by generating random parameters each time (Goulet, 2010). For example, the second time students worked on the same problem, they would need to plug in the new numbers/parameters generated into the correct formula in order to reach the right answer. This essentially decreased the possibility of guessing. In order to increase students' engagement in homework, a bonus system (Schilling, 2010) was implemented since 2010. In particular, students receive a 10% bonus for homework problems completed at least 24 hours in advance of the deadline. This bonus has resulted in over half of the homework being completed within the bonus time (Bennett et al., 2012). When an online homework system is used, detailed information about students' activities can be tracked and recorded into the access logs. For example, it is now possible to efficiently collect time stamps for each attempt towards a homework problem. As a result, researchers and educators are able to easily track students' learning process by looking at large volumes of data recorded by the online system.

The use of CMS and online homework enables researchers and educators to track and observe students' learning process by providing records of students' learning activities. However, few empirical studies focused on analyzing these types of student activities data, which includes huge amounts of information in multiple datasets. It is very difficult for traditional educational research approaches to deal with the vast quantities of data CMSs generated (Zorrilla, Menasalvas, Marin, Mora, & Segovia, 2005). Traditional educational research is hypothesis driven (Gaudioso & Talavera, 2006), which means that researchers start from a hypothesis and then they collect data to test the hypothesis. This approach would be effective when there are not too many variables and cases involved. However, it becaomes difficult when numerous factors and datasets are available to be analyzed. This challenge to the traditional educational research approach resulted in the rise of data mining in education course management systems (Romero, Ventura, & García, 2008).

In contrast to traditional research analysis, data mining is data driven, which helps explore the hidden pattern of large datasets. Minaei-Bidgoli et al. (2003) applied data mining techniques to predict students' final grades by analyzing students' learning activities data. They extracted variables of students' homework learning behavior such as the total number of correct answers, the total number of attempts, the number of problems that the student solves correctly on the first attempt, the time spent on a problem, and the total time spent on all homework, and conducted classification analysis. They transformed the target variable, final grades, into two classes (pass or fail), three classes (high, medium, and low), and nine classes (GPA divided into nine categories). Then they applied multiple classifiers, such as Quadratic Bayesian classifier, k-nearest neighbors, Parzen-window, multi-layer perceptron, and decision trees, to examine the significant predictors of the target variable. To improve the classifier performance, the authors used the genetic algorithm to combine multiple classifiers. With the combined model, the classification performance has reached 94.5% for the two classes prediction.

First Year Engineering Success/Retention Model

In order to combine different factors that influence first year engineering students' academic achievement, it is necessary to develop a working model of first year Engineering students' success in college study. There are a variety of first year retention or success model in the literature, but only a few were specifically developed for engineering students.

Veenstra et al. (2008) reviewed the literature and found that there were differences between non-STEM students and STEM students with regards to predictors of academic success and retention. For example, the SAT math score rather than SAT total score was more likely to predict STEM student success while SAT total score could predict non-STEM general college students' success better. They compared Pre-Med students, STEM students excluding engineering and Pre-Med students, non-STEM students and engineering students on nine categories of pre-college characteristics. Those nine categories included high school academic achievement, quantitative and analytical knowledge, study habits and independent learning, commitment to education and career goals, confidence in quantitative skills, commitment to enrolled college, financial needs, family support, and social engagement.

According to the group comparison results, significant differences were found in high school academic achievement, quantitative skills, and confidence in quantitative skills. Specifically, engineering students have higher ACT/SAT math, ACT/SAT science, and math placement scores. They also had higher self-rated math/science skills. Because of the differences found in engineering students and students majoring in other areas, the authors believed that it is necessary to have a model specifically developed for STEM or Engineering students. Therefore, they conducted regression analysis to develop a regression model of engineering students' success. Stepwise regression was used to determine the significant predictors of students' first year GPA in engineering studies. The result indicated that high school grades, quantitative and analytic skills, career goals, and confidence in quantitative skills were significant predictors of engineering students' first year GPA.

Later Veenstra et al. (2009) elaborated their first year engineering success model and built it into a first year retention model by adding a logistic regression model into the existing liner regression model. Their model was shown in figure 1. In this study, they also analyzed their data for the linear regression model and found the significant predictors of first year GPA as follow: high school grades, quantitative and analytic skills, career goals, and confidence in quantitative skills. They did not test the logistic model in this study.

There is no doubt that the series of study conducted by Veenstra, Dey, and Herrin (2008. 2009) helped identify the need to develop a first year success model for engineering students. Engineering students differed from their cohorts in other majors in a variety of aspects such as social engagement and high school performance. For example, the general non-STEM first year success model was insufficient in addressing the success of engineering students. Veenstra et al.'s regression models also pioneered the possible first year success models. However, there were some limitations that their study did not address. Firstly, from the perspective of psychometrics, there is an issue regarding the reliability and validity of the measurement instrument they used in this study. The CIRP freshman survey is a widely used survey and it has proven to be a reliable instrument (Keup, 2004). However, there is a lack of construct validity evidence indicating that items in the CIRP freshman survey could be divided into nine distinct categories. The authors conducted nine factor analyses in order to reduce multi-collinearity and to reduce the dimensionality of the predictors. However, it appears that there were no assumptions and statistics of factor analysis tests reported such as total variance explained by the factors and factor loadings of each item in the factors. There was no evidence reported regarding whether or not the factors were stable or the analyses were sufficient. In addition, there were numbers of factors that had only 2 items, which may indicate an unstable construct.

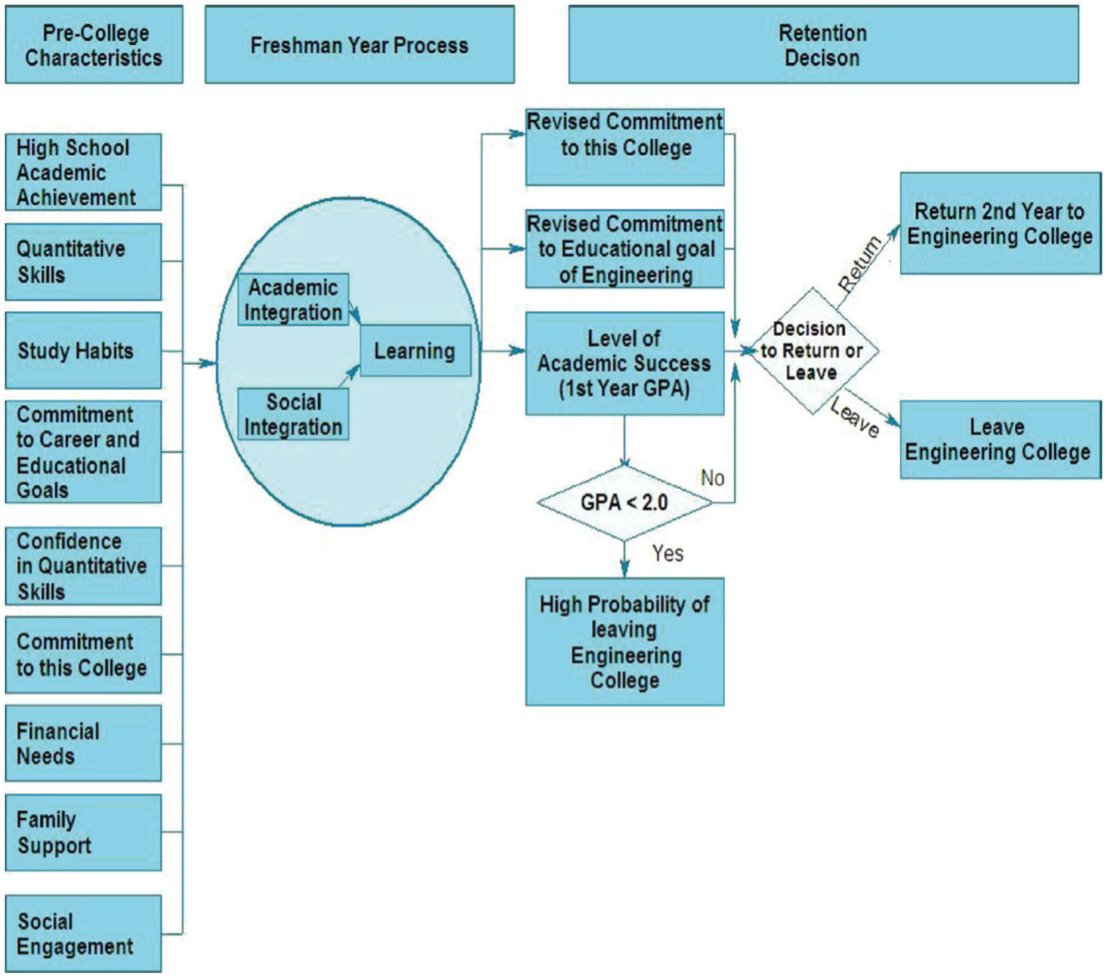

Figure 1. Engineering Retention Model by Veenstra, Dey, and Herrin (2009)

Secondly, Veenstra et al. (2009) indicated in their retention model that the first year success regression model is a “black-box” model ignoring students learning process in the first year. In their depiction of the model, although the pre-college characteristics could predict engineering students’ first year success (retention), the influences were mediated by students’ learning process. This means that students’ pre-college characteristics impact students’ learning process, and through students’ learning, student academic outcome is impacted. The linear regression model could not address the whole process. Summarily, the approach of applying factor analysis and linear regression appears to have resulted in the problems of low validity, and it did not address students’ learning process at their first year. Therefore, another more appropriate approach is needed to address the above problems.

Psychometric Properties of Instruments

Researchers need to follow appropriate steps to ensure that a survey instrument has acceptable reliability and validity. It is necessary to develop a reliable and valid instrument to investigate engineering students’ pre-college characteristics. Reliability is defined as the extent to which a test/scale can get consistent or similar result across the time (Furr & Bacharach, 2014). There are different ways we could measure reliability.

Test-retest reliability: Involves administering the same test twice to the same group after a certain time interval has elapsed. Reliability is calculated by the correlation between two sets of data. This is an effective way to measure reliability, but the problem is that we cannot always get participants to fill out the survey in certain periods of time on two occasions. Also because of the potential practice effect, researchers should be aware of the time interval between two administrations of the survey. If the interval is too short, participant could remember the items. If the interval is too long, some internal change will occur within the participants.

Equivalent form reliability: Two different but equivalent (alternate or parallel) forms of an instrument are administered to the same group during the same time period. This one helps with test-retest method’s disadvantage. We do not need our participants to fill out the survey later again. But it will be a challenge to develop two equivalent forms.

Internal-consistency methods: include calculating cronbach’s alpha and split half procedure. This is based on the degree of association among the items in the scale/sub-scale. Cronbach’s alpha is ranging from 0-1 with higher value represent higher internal reliability.

Validity is defined as the extent to which the instrument measures what it intends to measure (Furr & Bacharach, 2014). There were three major types of validity:

Translation validity: which included face validity and content validity, criteria validity, and construct validity. Face validity is the weakest way to measure validity. To judge whether or not the test has face validity, researchers simply read the items and subjectively judge whether or not the items measure the construct. Content validity addresses the match between test questions and the content they are intended to measure.

Criteria validity also included two types of validity: predictive validity and concurrent validity. Predictive validity compares the instrument scores and criterion scores at a later time while concurrent validity compare instrument data and criterion data at the same time.

Construct validity is related to how well the instrument measures the theoretical constructs it intends to measure.

Table 1 summarized different aspects of psychological properties such as reliability and validity.

Data Mining within a Course Management System

After making sure that the survey instrument used meets the acceptable levels at different psychometrics properties, we need to choose appropriate statistical model. Decision tree is a widely used classification approach in the education data mining area (Romero, Ventura, & García, 2008). It is popular because of its flexibility and interpretability (Huo, Kim, Tsui, & Wang, 2006; Moon, Kang, Jitpitaklert, & Kim, 2012). Although the outcome variable in a decision tree is usually restricted to a binary variable, both categorical and continuous predictors can be handled well in a decision tree. There are multiple purposes for the application of decision tree analysis in educational research. First of all, it is similar to logistic regression that multiple factors were used to predict students' performance and final grade (Minaei-Bidgoli & Punch, 2003). The advantage of decision tree over logistic regression model is its flexibility to handle both continuous and categorical variables and interpretability (Huo, Kim, Tsui, & Wang, 2006; Moon, Kang, Jitpitaklert, & Kim, 2012). Moreover, decision tree can be used to group students who have similar characteristics and learning patterns by generating segment rules (Chen, Liu, Ou, & Liu, 2000).

Conclusion

This chapter provided a synthesis and critique of the theoretical and empirical literature on the factors related to engineering students' learning outcomes, related engineering student success models, and appropriate approaches to develop a working model to predict engineering students' learning outcomes.

Table 1.

Psychometrics Properties of Survey Instrument (Furr & Bacharach, 2014)

Domain	**Psychometric Property**	**Aspect of a psychometric property**	**Definition**
Reliability	*Test-retest reliability*		Test-retest reliability involves administering the same test twice to the same group after a certain time interval has elapsed. Reliability is calculated by the correlation between two sets of data.
	Internal consistency		Internal consistency is measured by correlation among items with a scale/subscale.
	Equivalent form reliability		Equivalent form reliability is measured by having different but equivalent (alternate or parallel) forms of an instrument administered to the same group during the same time period.
Validity	Translation validity	Face validity	To judge whether or not the test has face validity, just read the items and subjectively judge whether or not the items measure the construct.
	Translation validity	Content validity	Content validity addresses the match between test questions and the content they are intended to measure.
	Criteria validity	Predict validity	Predictive validity compares the instrument scores and criterion scores at a later time
	Criteria validity	Concurrent validity	Concurrent validity compare instrument data and criterion data at the same time.
	Construct validity		Construct validity addresses how well the instrument/experiment measures the underline theoretical construct it aim to measure.

Engineering education in higher education is currently at a crossroad. According to the report, Engineers of 2020, future engineers are expected to have strong analytic skills, practical ingenuity, creativity, communication, principles of business and management, leadership, sense of professionalism, dynamism, agility, resilience, flexibility, and become a lifelong learner (Clough, 2004). Engineering education assumes the responsibility to cultivate increasing numbers of professional engineers equipped with essential engineering skillsets. However, engineering students' retention rates in higher education are decreasing (Ohland et al., 2008), which resulted in an unmet market demand on engineering jobs (Weiss, 2009).

In order to increase the retention rate of engineering students and better train engineering students to master the necessary skillsets, Engineering researchers and educators conducted numerous empirical studies to identify factors that related to engineering students' retention. Some of the empirical studies (e.g., Veenstra et al., 2008; Ahmad et at., 2012) put emphasis on pre-college characteristics that influence engineering students' first year academic success. According to the literature, these pre-college characteristics included: high school performance, self-efficacy on quantitative and science skills and knowledge, motivation, knowledge about engineering professions and education, engineering study skills, and students' social/family background.

Other empirical studies (Bernold, 2007; Liberatore, 2011) focused on improvements of engineering curriculum and teaching approaches to better engage first year engineering students. For example, Bernold proposed that inquiry-based learning is "the pedagogical paradigm for 21st century" (Bernold, 2007). Compared to the traditional lectures that the professors spend most of the time filling students' "empty brain", the inquiry-based learning requires students to manage their own learning. Students need to know the way of acquiring knowledge, developing personal strategies, recognizing their personal strength and weakness, and gaining new knowledge.

The University of Tennessee *Engage Program* incorporated inquiry-based and project-based teaching into the traditional lectures. The program courses use a customized web-based homework system (Schleter & Bennett, 2006). This system provides personalized homework by generating random parameters each time (Goulet, 2010). In order to increase students' engagement in homework, a bonus system (Schilling, 2010) was implemented since 2010. In particular, students receive a 10% bonus for homework problems completed at least 24 hours in advance of the deadline. This bonus system has resulted in over half of the homework being completed within the bonus time (Bennett et al, 2012). When an online homework system is used, detailed information about students' activities can be tracked and recorded into the access logs. For example, it is now very easy to collect time stamps of each attempt towards a homework problem. As a result, researchers and educators can easily track students' learning process by looking at large volume of data recorded by the online system.

It is necessary to consider both pre-college characteristics and students' first year learning process when examining factors influence students' first year success (need citations here). Accordingly, a theoretical framework that encompasses both pre-college characteristics and students' first year learning process is needed. However, there are few studies focusing on developing a first year success model for engineering students. Although there are large amount of first year success models existing in the literature, such as Tinto's interactionalist theory model (Tinto, 1975, 1993) and Astin's theory of involvement model (Astin, 1984), these establisehd models were not specifically developed for engineering students, and the literature suggests that engineering students are different. For example, Veenstra, Dey, and Herrin (2008) compared engineering students with students in other STEM areas, students in non-STEM area, and Pre-Med students, and found that Engineering students differed from their cohorts in other disciplines in varies aspects such as quantitative skills and social engagement. They stated that models in the literature developed using first year student without taking account disciplinary differences can not apply to engineering students directly. There is a need to develop a specific first year engineering student success model.

Veenstra, Dey, and Herrin (2008) developed a first year engineering success model using the Cooperative Institutional Research Program (CIRP) freshman survey. The survey was administered nationally by the Higher Education Research Institute (HERI) at UCLA's Graduate School of Education and Information Studies. They arbitrarily divided the survey items into nine factors/pillars then conducted 9 factor analyses to identify constructs under each factor. They developed a regression model using the factor scores to predict students' academic success. However, the study failed to provide reliability and validity information of the instruments. Although CIRP freshman survey was a widely used survey instruments, there was no supportive validity evidence in the study indicating the survey could be divided into 9 aspects. This study did not include students' first year learning process variables in the model although they claimed students learning process were also important factor of engineering students' first year success at their later publication (Veenstra, Dey, & Herrin, 2009).

The use of online course management system (CMS) and online homework enables researchers to collecting objective data about students' homework learning process. For example, Online CMSs keep track of students' online activity logs, such as logins, content contributions (e.g., creating a forum post), and homework attempts ((Romero, Ventura, & García, 2008; Minaei-Bidgoli, Kashy, Kortemeyer, & Punch, 2003). Applying data mining techniques, researchers could extract student level feature variables such as how long student spent on homework, whether or not student get homework question answered correct, when did students start to work each homework questions, and how many attempts students had on each homework questions, etc.

In summary, in order to develop a working model that accurately predicts engineering students' academic success/fail at early stage, it is necessary to make sure all of the factors such as pre-college characteristics and college learning activity attributes are included in the model. It is also essential to make sure that the survey instrument is reliable and valid to collect information about students' pre-college characteristics. Lastly, decision tree, as a popular prediction model in data mining, is proposed as the appropriate statistical model to address the research questions.

Chapter 3

Method

Chapter 3 introduces the research design and methodology of the study. Firstly, the chapter briefly reviewed the research purpose of the study. Then this chapter discussed the sampling methodology, participants' demographics, development of the data collection instrument for the study, measures that were used, and study procedures.

Purpose of the Study

In the present study, the author investigated students' pre-college characteristics such as exposures to engineering education and professionals, and psychological traits such as self-confidence in math and science skills, motivation of studying in Engineering. The author also applied data mining techniques to describe students' engineering homework learning activities. Framed by Tinto's interactionalist theory model (1993) and Veenstra et al.'s first year engineering retention model (2009), the author aimed to create a working model to predict students at risk in an Introduction to Engineering course by considering both students' pre-college characteristics, psychological traits, and online homework learning behavior. In this way, this study can help the course instructor to create an early warning system and develop targeted interventions for students at risk.

Hypothesis

The Engineering Students' Pre-college Characteristics survey demonstrates an acceptable internal consistency reliability (Cronbach's alpha >.60) and validity (construct).

Research Questions

1. How actively are first year engineering students engaged in their online homework study?
2. Are there any group differences (gender and first generation) on students' pre-college

characteristics, psychological traits, and online homework activities?

3. What factors are associated with students' risk to fail a course?
4. What are the characteristics of students at risk?

Research Design

This study used a cross sectional design, which collects survey and observation data from a large number of individuals (Colton & Covert, 2007). One part of the data was collected through an online survey, which consisted of items to assess students' precollege characteristics including demographic information, pre-exposure to engineering professionals, and psychological traits such as motivation in studying in Engineering. In addition to the survey data, the study also collected observation data about students' homework learning activities. This part of data was collected through mining the engineering online course management system logs.

Population and Sampling

The population of interest was first year engineering students in the United States (US). Participants for the study were recruited through purposive sampling approaches. First year engineering students enrolled in the Introduction to Engineering (EF151) course at the University of Tennessee, Knoxville, were selected as the study sample.

Participants

First year engineering students enrolled in EF151 participated in the study. Because the separate study consent students in the online survey and the release of online homework activity log, the number of participants enrolled in two data collection parts were different (refer to data collection process below for more information). The responses of survey participants were included in the analysis of survey reliability and validity. Among the 585 students, 408 students responded to the survey, the response rate was 69.7%. Only students who agree to participate in both the survey and release their online homework activity log data were included in the group comparison and decision tree model building. There were 291 out of 585 students agreed to both respond to the survey and release their only homework logs. The response rate was 49.7%.

After the data cleaning process, 392 out of 408 first year engineering students remained as survey participants. Detailed demographic information was shown on Table 2. Among the 392 participants, 72.4% are male students and 26.8% are females. The majority of participants (82.7%) are Caucasian. Fifty-six out of 392 students are the first generation in their family to go to college. Only 6 students are part-time students. There are 117 students (29.8%) who have immediate family members (parents or siblings) holding an engineering degree. There are 145 out of 392 students (37.0%) reporting that they never took courses related to engineering before college.

Only 291 students agreed to both release their online homework activity logs and participated in the survey (detailed demographic information was shown on Table 2). Overall, the resulting sample structure is similar to the survey participants' sample. Among the 291 participants, 72.5% are male students and 27.1 % were females. The majority of participants (83.2%) are Caucasian. Forty-four out of 291 students are first generation in the family to go to college. Only 6 students are part-time students. There are 87 students (29.9%) who have immediate family members (parents or siblings) holding an engineering degree. There are 109 out of 291 students (37.5%) reporting that they never took courses related to engineering before college.

Data collection process

There were two parts of data collected from the students: 1) self-reported online survey data, and 2) students' homework learning activity data extracted from CMS.

Survey data collection procedure.

First year engineering students were recruited at the beginning of 2015 Fall semester to take the Engineering Pre-college Characteristics Survey. At the beginning of the semester, we drafted the recruitment email and sent it to the course instructor. The incentive, which is extra credit in the class, the research purpose, and the request for consent to participate were all indicated in the email. The course instructor sent the recruitment emails with the link to the survey in the email to the students. A reminder email was sent eight days later to recruit more students.

Data mining process in CMS.

The data mining process started with writing a data requesting letter (See appendix1for data description table). Upon IRB approval, I sent the data-requesting letter to the course instructor and the data management staff. It took the data management staff 2-3 weeks to prepare and send the data in the format of Comma Separated Values (.CSV). The feature extraction process started as soon as the datasets were received. The process included data aggregation and variable computation. In particular, as listed in Figure 2, each box in the figure represents a dataset, with variables listed under the header. The HWSets dataset is the description about each homework set.

The HWperform datasets is the final datasets containing all of the extracted variables. Those variables are described as following

Pearly, Pregular, Plate: percentage of homework assignments accomplished during bonus period, regular period (i.e., the day when the homework is due), and after deadline (i.e., late).

PansweredQ: percentage of homework problems the student has answered. This variable is calculated by dividing the number of problems a student has had attempts on by the total number of available problems in the homework assignment.

Mattempts: the average number of attempts per problem. This variable is calculated by dividing the total number of a student's attempts by the total number of problems.

Correctness: percentage of problems solved correctly. This variable is calculated as the total number of a student's correct answers divided by the total number of problems.

LearningSpeed: the average number of attempts per correct answer. The LearningSpeed variable may be interpreted as the number of attempts a student needs on average to find the correct answer. It is calculated by dividing the total number of attempts of a student by the total of correct answers.

Preparation: the percentage of problems a student solves correctly on the first attempt. For each of the homework problem, if a student can answer it correctly without hint or incorrect attempts, the course instructor believes the student is well prepared for the homework. A student may be prepared by in-class learning through lectures, or collaborative-learning through group study.

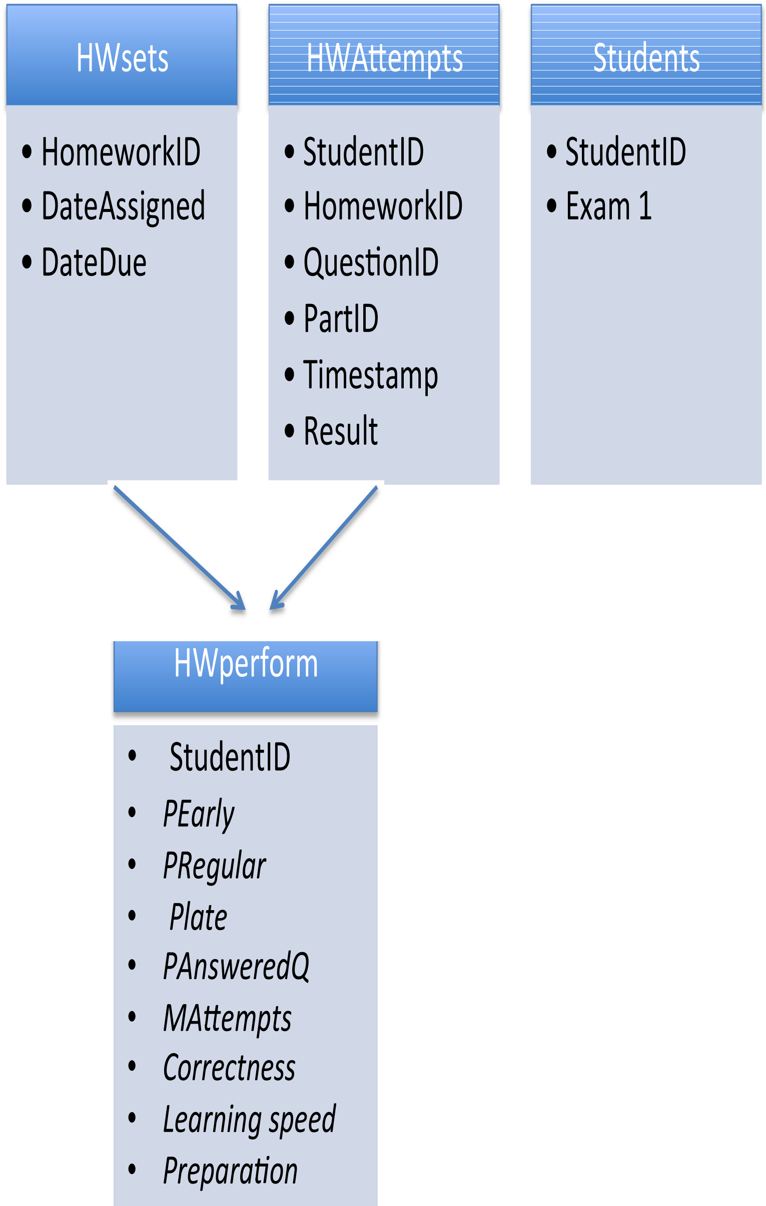

Figure 2. The Process of Data Mining

Table 2.

Participants' Demographics

Variable	Survey Participants		Model Building Participants	
	N	%	N	%
Gender				
Female	105	26.8	79	27.1
Male	284	72.4	21	72.5
Prefer not to answer	3	0.8	1	0.3
Ethnicity				
American Indian	1	0.3	1	0.3
Asian	23	5.9	15	5.2
African American	9	2.3	6	2.1
Multiracial	9	2.3	8	2.7
Caucasian	324	82.7	242	83.2
Latino	7	1.8	4	1.4
Other	12	3.1	10	3.4
Prefer not to answer	7	1.8	5	1.7
First Generation				
Yes	56	14.3	44	15.1
No	336	85.7	247	84.9
Full-Time Student				
Yes	385	98.2	286	98.3
No	6	1.5	5	1.7
Prefer not to answer	1	0.3	0	0.0
Family member has engineering degree				
Yes	117	29.8	87	29.9
No	271	69.1	201	69.1
Prefer not to answer	3	0.8	2	0.7
Classes related to Engineering				
0	145	37	109	37.5
1	70	17.9	55	18.9
2	74	18.9	52	17.9
3	48	12.2	37	12.7
4 or more	55	14	38	13.1
Total	392	100	291	100

Survey Instrument

The process of creating survey items began with reviewing literature related to engineering students' pre-college characteristics. After identifying the 6 aspects of pre-college characteristics (self-efficacy on math and science skills and knowledge, motivation, knowledge about engineering professions and education, engineering study skills, and students' social/family background), I adapted survey items from published survey instruments: the Academic Pathways of People Learning Engineering Survey (APPLES). APPLES is a widely used survey instruments with acceptable reliability (Sheppard et al., 2010). Specifically, items related to motivation, knowledge about engineering professions and education, engineering study skills, and students' self-efficacy on quantitative and science skills and knowledge were adapted from APPLES, and students' social/family background items were developed by the researchers.

After writing the items, I used an online technological tool (http://read-able.com) to determine the reading level of the items and investigated their wording by Microsoft WORD spelling and grammar tools. I sought to use simple and readable language to make sure that the first year engineering students could understand my questions. The reading level is 5.5 grades, which suggests that the survey items are easy for first year engineering students to understand. The drafted survey was then reviewed by survey research experts, who are professors and PhD students in Evaluation, Measurement, and Statistics program at the University of Tennessee. All of the experts have a wealth of experience in survey development. They provided beneficial feedback to revise my drafted survey. After modifying the survey according to the survey experts' feedback, I created the online version of the survey using Qualtrics. Links were sent to the survey experts again to pilot test the survey. The final version of the Engineering Students' Pre-college Characteristics Survey (ESPCS) was created by revising the survey according to the pretest feedbacks. The survey has 44 questions in total. The 44 questions are divided into 5 parts:

Demographic Information: There are 9 questions in this section asking about students demographic and family information. Sample questions are “what is the highest education level of your mother/father?” and “what is your ethnicity?”

Students’ engineering study skills: there are 11 items in this part that are adapted from APPLES. Participants are required to rate statements on a 5-point Likert scale with 0 representing "never" and 4 representing "always". Sample questions include, “Evaluate the quality or reliability of information you received” and “Seek alternative solutions to a problem”.

Motivation: there are 16 items in this part that are adapted from APPLES. Participants are required to rate the statement on a 5 point Likert scale with 1 representing Strongly disagree and 5 representing strongly agree. Sample questions are: “My parent(s) would disapprove if I chose a major other than engineering” and “Engineers have contributed greatly to fixing problems in the world”.

Self-efficacy on Math and Science Skills: there are 4 items in this part that are adapted from APPLES. Participants are required to estimate their own competence compared to their classmates on a 5-point Likert scale with 1 representing lowest 25% and 5 representing highest 25 percent. Sample questions are: “Math ability” and “Ability to apply math principles in solving real world problems”.

Knowledge about Engineering Professions: there are 5 items in this part that are either adapted from APPLES or brainstorming. Sample questions are: “How many classes related to Engineering did you take before college?” and “Before college, how much knowledge did you have about the engineering profession?”

To ensure the safety and privacy of the participants, all ethical guidelines were followed as outlined in the Ethical Principles of Psychologists and Code of Conduct (http://www.apa.org/ethics/code/principles.pdf) published by American Psychological Association in 2010. Two separate consent forms were developed for the two parts of data collection processes. There are a couple of reasons to develop multiple consent forms. First of all, it was required by the Institutional Review Board (IRB) to obtain written permission from students when researchers are requesting data protected by Family Educational Rights and Privacy Act (FERPA). Secondly, giving the experience of the course instructor, students are more willing to participate in a survey study while they may refuse to release their online homework learning activity logs. In addition, although conducting CFA requires relatively strict sample sizes (N > 300), decision tree, as a more descriptive approach, does not require large sample size. In summary, separate consents did not impact the assumptions of the study techniques.

Data were collected through a password-protected account on the Qualtrics online survey software, and only the principal investigator of the study had the access to the account and survey data. Identities of the participants were moved from the downloaded data. Instead, the researchers created a research-specific ID for each participant. All collected data were kept confidential.

Building Decision Tree Model

After survey and online homework activity logs were collected, I applied decision tree algorithm to build a partition based first year engineering student success model (Romero, Ventura, & García, 2008). Decision tree is a tree-like flow chart with parent (root) nodes and child (leaf) nodes. As illustrated in Figure 3, all parent nodes must have 2 or more child nodes. There are splits for each parent node, where certain cut-offs of the attributes/factors are used to group the data into two or more child nodes. The interpretation of the decision tree is also very straightforward because the flow chart of decision tree is an illustration of "if then" algorithm. For example, as shown in Figure 3, if a student's attribute1 is larger than the cut off, X, he will be in parent node 1, then he will be in child node 1; branching from parent node 1, if a student's attribute2 is larger than the cut off value, Y, he will be categorized into child node3, then he will be in child node2. Using a specific example in education settings, the target is student pass/fail a course. Attribute 1 is students' average daily study hours with the cut off as 2 hours. If a student studies more than 2 hours a day, he will be put in parent node 1 and have higher possibility than other students to pass a course. Attribute 2 is students' in class quiz score and the cut off is 80 out of 100. If a student studies more than 2 hours a day and scores over 80 in in-class quiz, he will be nested in child node 3 and have a higher possibility to pass the course than students in child node 3, whose in class quiz score is below 80 even though they study more than 2 hours a day.

Decision tree is popular because of its flexibility and interpretability (Huo, Kim, Tsui, & Wang, 2006; Moon, Kang, Jitpitaklert, & Kim, 2012). One of the advantages of decision tree model over regression model is that it can divide students into different groups based on their characteristics (Chen, Liu, Ou, & Liu, 2000). This trait of decision tree is similar to cluster analysis, which is a popular clustering technique to identify groups of participants with similar behavior patterns (Romero, Ventura, & García, 2008). However, different from cluster analysis, which is more of a descriptive analysis, decision tree classify individuals based on both the identified predictors of the target variable and the relationships between predictors and target variables (i.e., importance of predictors, the value of target variables).

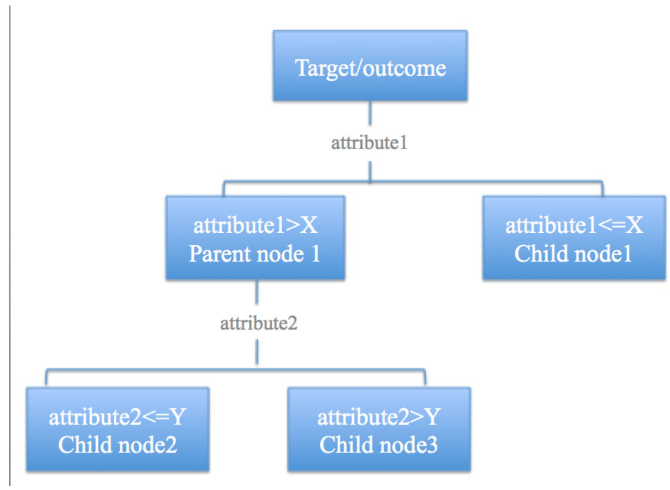

Figure 3. Example of A Decision Tree

Chapter 4

Results

The present study investigated students' pre-college characteristics such as exposures to engineering education and professionals, and psychological traits such as self-confidence in math and science skills, and motivation of studying in Engineering. It also enabled the application of data mining techniques to describe students' engineering homework learning activities. Framed by Tinto's interactionalist theory model (1993) and Veenstra et al.'s first year engineering retention model (2009), the study aimed to create a working model to predict students at risk academically in an introduction to Engineering course by considering students' pre-college characteristics, psychological traits, and online homework learning behavior. In this way, the study may be possible to help the course instructors seeking to create an early warning system and develop targeted interventions for students at risk.

The hypothesis was that the Engineering Students' Pre-college Characteristics survey demonstrates an acceptable internal consistency reliability (Cronbach's alpha >.60) and validity (construct). The research questions addressed were as follows:

1. How actively are those first year engineering students engaged in their online homework study?
2. Are there any group differences (gender and first generation) on students' pre-college characteristics, psychological traits, and online homework activities?
3. What factors are associated with students' risk to fail the course?
4. What are the characteristics of students at risk?

Data Cleaning

Data cleaning followed the Twelve Steps of Data Cleaning by Morrow and Skolits (Morrow & Skolits, 2014), and assessed for the assumptions of Confirmatory Factor Analysis (CFA). Firstly, frequency analyses were conducted on each of the variables and coding errors were corrected. One student reported the age as 17, which is outside of my recruitment sample. All data related to this student were removed from the dataset.

Next, the data were analyzed both case-wise and variable-wise for missing values. Only the records that had more than 80% of the survey questions answered were kept for further analysis. 16 out of 408 cases were deleted due to incompletion of survey. After case-wise missing value analysis, 392 cases remained in the dataset. In the variables that had missing values, the amount of missing data was less than 1% per variable except for the variable ms3. Accordingly, 5% or fewer amounts of random missing values in a large dataset is considered acceptable (Tabachnick, & Fidell, 2013). The variable math and science self-efficacy item 3(ms3) was removed from analysis because more than 50% of the students did not respond to the variable. The missing values in the survey were not replaced, and listwise deletion was chosen as the method for handling missing data during analyses (Allison, 2001).

Then I ran s second round frequencies and descriptive statistics to search for outliers and deal with the outliers. This included all of the Likert-scale items in the survey including items related to students' motivation in studying in Engineering, study habits, and self-efficiency in math and science skills. If a score is more than 3.29 standard deviations above or below the mean, it was considered an outlier (Tabachnick & Fidell, 2013). Since deleting the outliers didn't lead to too many missing data (missing data > 5%), the outliers were recoded into missing data.

Univariate normality of the self-efficacy variables was assessed by skewness and kurtosis values. According to Tabachnick and Fidell (2013), Skewness and kurtosis values less than |2| indicate that the variables are normal distributed. All of the variables in the dataset have Skewness and kurtosis values less than |2|, thus the univariate normality was assumed.

Hypothesis: Survey Validation

Structural equation modeling (SEM) approach was applied to assess the instruments' construct validity and reliability (Kline, 2005). Structural equation modeling (SEM) is a causal model in which the paths in a graphic model are expressed as a series of algebraic equations (Boyd, Frey, & Aaronson, 1988). When applying SEM to assess the construct validity of an instrument, the study aimed to:

1.Test the hypothesized factors;

2. Assess whether or not each item significantly loads in the construct; and

3. Examine whether or not the constructs are interdependent from each other (LaNasa, Cabrera, & Trangsrud, 2009).

In a structural equation model, measureable and observed variables are known as manifest variables that are predictors of construct. Those observed/manifest variables are represented by squares in the graphic SEM. The constructs, which are not observable but exist in the model are called latent variables. Latent variables are represented by circles in the graphic SEM. In my study, all of the survey items as well as extracted feature variables described above are manifest variables. All of the constructs predicted by the observed variables such as factors in the survey are considered as latent variables. Model testing section has two parts: measurement model testing (confirmatory factor analysis, or CFA) and structural model testing (path analysis). Since the purpose of applying SEM in this study was to examine the validity of the instrument, only the measurement model testing part of SEM (CFA) was conducted. The model testing analysis was conducted using SPSS AMOS 22.0. Chi-square, root mean square error of approximation (RMSEA), Goodness-of-fit statistic (GFI), the adjusted goodness-of-fit statistic (AGFI), and CFI (Comparative fit index) were reported as the model fit indices.

First of all, a confirmatory factor analysis was conducted according to the model shown in figure 4. The initial model was based on the APPLES survey constructs. When fitting the model, we modified the model by adding covariates between constructs and within-construct manifest variables.

When considering the fitting index of the model fit, chi-square value was first applied. However, the chi-square value is sensitive to sample size. When the sample size increased, the Chi-square value also increased. Thus the Chi-square of large sample size will have an artificial tendency to reject the model (Dimitrov, 2008). Because the Chi-square value could not provide sufficient and valid evidence for model fit (Bentler & Bonett, 1980; Raykov & Marcoulides, 2006), other indices were used. Specifically, chi-square to degrees of freedom ratio, the goodness-of-fit index (GFI), Comparative fit index (CFI), root mean square error of approximation (RMSEA) index were applied in the present study. As Bollen (1989) suggested, a Chi-square to degrees of freedom ratio smaller than 2.00 indicated an acceptable model fit. Kline suggests that the value of chi-square to degrees of freedom ratio smaller than 3 signifies a good fit of the model (Kline, 2005). This model had the value of 1.91, which indicated an acceptable model fit. However, the result of other fitting indices showed that the hypothesized structure model did not fit well. As can be seen in table 3, the goodness of fit index was 0.85, and the CFI index was 0.84, both of them were lower than 0.90. Generally, a comparative fit index (CFI) greater than .93 indicated a good model fit though a value greater than .85 reveals somewhat acceptable fit (Hu & Bentley, 1999). In regards of goodness-of-fit (GFI) index, a value greater than .95 is considered as a reasonable level while the value that is higher than .90 could be somewhat acceptable model fit (Hu & Bentler, 1999). Hu and Bentler (1999) suggested 0.06 as the cut-off value of RMSEA to indicate a accept model fit. Any RMSEA values smaller than 0.06 indicate good model of fit.

Table 4 shows the item loadings on each construct. As indicated by the item loadings, mentor and parent motivation are not valid constructs. Mentor motivation is not a significant indicator of the general motivation construct. Neither of these survey items significantly loads on mentor motivation. There are only two items in the parent motivation construct and the loadings are not significant. Therefore, both the item loadings and model fit index indicate that the hypothesized survey instrument do not have acceptable construct validity.

Table 3

Fitting Indices of Hypothesized Model

χ^2	*df*	χ^2/df	*GFI*	*CFI*	*RMSEA*
1039.38	545	1.91	0.85	0.84	0.05

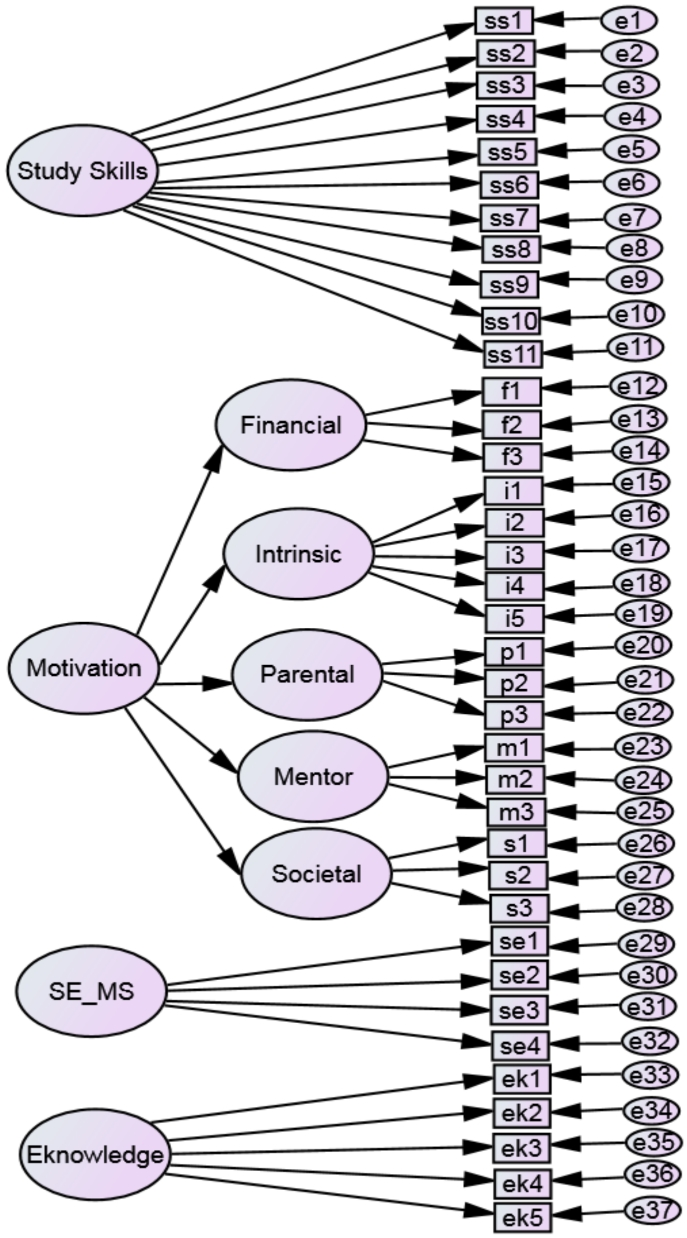

Figure 4. Hypothesized Instrument Structure

Note. SE_MS is self-efficacy in math and science. Eknowledge is knowledge about engineering professions and education.

Table 4

Item Loadings of The hypothesized model

			Loadings	S.E.	C.R.	P
Financial	<---	Motivation	1.000			
Intrinsic	<---	Motivation	4.433	1.837	2.414	.016
Parental	<---	Motivation	-2.247	1.047	-2.146	.032
Mentor	<---	Motivation	-.003	.658	-.005	.996
Societal	<---	Motivation	2.614	1.120	2.334	.020
f2	<---	Financial	1.478	.464	3.185	.001
f3	<---	Financial	.671	.653	1.028	.304
i2	<---	Intrinsic	1.213	.145	8.394	***
i4	<---	Intrinsic	1.244	.123	10.118	***
i5	<---	Intrinsic	1.069	.114	9.362	***
s1	<---	Social	1.000			
s2	<---	Social	1.151	.204	5.647	***
ms2	<---	SE_MS	2.164	.272	7.961	***
ms4	<---	SE_MS	1.553	.165	9.409	***
ms1	<---	SE_MS	1.000			
s3	<---	Social	1.715	.261	6.560	***
m2	<---	Mentor	-143.576	28252.804	-.005	.996
m1	<---	Mentor	1.000			
m3	<---	Mentor	-142.872	28116.522	-.005	.996
p1	<---	Parental	1.000			
p2	<---	Parental	-.454	.319	-1.421	.155
i3	<---	Intrinsic	1.278	.106	12.042	***
i1	<---	Intrinsic	1.000			
f1	<---	Financial	1.000			
ss1	<---	Study_Skills	1.000			
ss2	<---	Study_Skills	1.296	.285	4.552	***
ss3	<---	Study_Skills	1.342	.291	4.616	***
ss4	<---	Study_Skills	1.143	.286	3.998	***
ss5	<---	Study_Skills	1.358	.303	4.482	***
ss6	<---	Study_Skills	1.074	.263	4.085	***
ss7	<---	Study_Skills	1.491	.318	4.687	***
ss8	<---	Study_Skills	1.814	.390	4.656	***
ss9	<---	Study_Skills	1.639	.362	4.527	***
ss10	<---	Study_Skills	1.322	.302	4.371	***
ss11	<---	Study_Skills	1.361	.315	4.317	***
ek1	<---	Eknowledge	1.000			
ek2	<---	Eknowledge	1.272	.280	4.543	***
ek4	<---	Eknowledge	.334	.085	3.924	***
ek3	<---	Eknowledge	.293	.082	3.576	***
ek5	<---	Eknowledge	.380	.088	4.328	***

An alternative model was built by deleting the parent and mentor constructs. The new model led to item loadings as shown in Figure 5. All of the factor loadings in the new model are significant. The model fitting indices were summarized in Table 5. According to the model fitting index cut-off suggested by previous researchers (Hu & Bentler, 1999), the modified model well fits the data and thus indicates a valid structure.

All of the factor loadings are significant as showed in Table 6, which also indicates a valid structure of the model. In order to test internal reliability of the scales, all of the Cronbach's alpha values were calculated and listed in Table 6. Only the scale of engineering knowledge shows a poor internal reliability and all other scales have acceptable internal reliability. In summary, the modified instrument used in the present study has acceptable internal reliability and construct validity.

Table 5

Fitting Indices of Modified Model

χ^2	*df*	χ^2/df	*GFI*	*CFI*	*RMSEA*
577.56	364	1.59	0.90	0.92	0.04

Table 6
Standardized Item Loadings of the New Model

Measure		Loading	Reliability
Study Skills			
ss1	Ask questions in class	.31***	
ss2	Support your opinions with a logical argument	.49***	
ss3	Seek solutions to problems and explain them to others	.51***	
ss4	Revise your papers to improve your writing	.32***	
ss5	Evaluate the quality or reliability of information you received	.47***	
ss6	Take a risk because you feel you have more to gain	.36***	.75
ss7	Seek alternative solutions to a problem	.54***	
ss8	Look up scientific research articles and resources	.55***	
ss9	Explore topics on your own, even though it was not required for a class	.52***	
ss10	Accept mistakes as part of the learning process	.43***	
ss11	Seek feedback on your academic work	.40***	
Intrinsic Motivation			
i1	I feel good when I am doing engineering	.52***	
i2	I like to build stuff	.64***	
i3	I think engineering is fun	.69***	.81
i4	I think engineering is interesting	.88***	
i5	I like to figure out how things work	.69***	
Financial Motivation			
f1	Engineers make more money than most other professionals	.11***	.78
f2	Engineers are well paid	.18***	
Societal Motivation			
s1	Technology plays an important role in solving society's problems	.33***	
s2	Engineers have contributed greatly to fixing problems in the world	.43***	.69
s3	Engineering skills can be used for the good of society	.68***	
Self-efficacy in Math and Science			
ms1	Math ability	.51***	
ms2	Science ability	.98***	.77
ms4	Ability to apply science principles in solving real world problems	.75***	
Engineering Exposure before College			
ek1	Knowledge about engineering professions before college	.52***	
ek2	Exposure to a professional engineering environment as a visitor	.82***	
ek3	Exposure to a professional engineering environment as an intern	.24***	.57
ek4	Exposure to a professional engineering environment as an employee	.27***	
ek5	Family members (parents, siblings) hold an engineering degree	.31***	

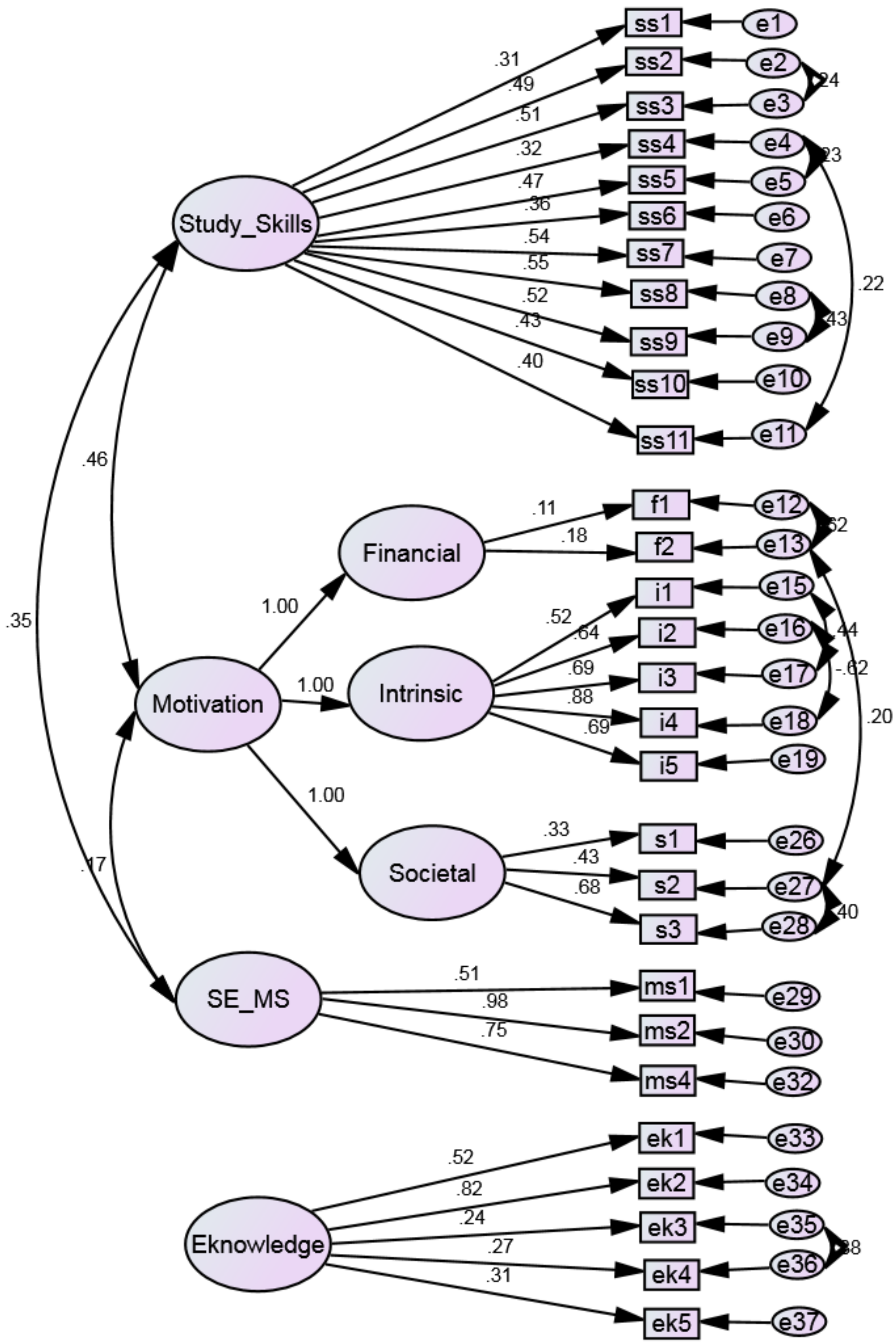

Figure 5. Modified Survey Instrument Structure

Research Question 1: How actively are first year engineering students engaged in their online homework study?

In order to answer this research question, descriptive analyses were run using the extracted online homework activities variables. Mean and standard deviation of each variable were listed in Table 7. Students were actively engaged in their online homework study in several different ways. First of all, the majority of their homework attempts were recorded in the bonus time, which indicates that students made use of the early completion bonus by starting work on their homework one day earlier than the deadline. Secondly, students have attempts on the majority of their homework questions (97.5%) no matter they answered it correctly or not. Thirdly, students have more than one attempt on their homework questions and they have relatively high correct rate (96.4%).

Research Question 2: Are there any group differences (gender and first generation) on students' pre-college characteristics, psychological traits, and online homework activities?

Independent samples t-tests were conducted to compare pre-college characteristics, psychological traits, and online homework activities features between male and female students. Levene's test for equality of variances was not significant (p = .18) for pre-college characteristics, psychological traits, and average attempts (Mattempts) indicating that population variances of those variables for both groups were equal. However, Levene's test was significant (p < .001) for the majority of homework activity features, and therefore the t-test results for samples with unequal variances was used to compare the average scores of these variables between male and female students. Table 7 showed the mean, standard deviation of the variables, as well as the t-tests comparison results. Male and Female students did not differ in their pre-college characteristics. However, female students were more actively engaged in their online homework studies. They have more homework attempts one day earlier than the deadline and they are less likely to solve the homework problem after deadline. Secondly, they have more attempts than male students on homework. Finally, the correct rate of female students is higher than male students.

Similar procedures were followed to examine the first generation students and non-first generation students on all of the variables. The results were also summarized in Table 7. Students who are the first generation in their family to go to college have less exposure to engineering professionals/educations before college than their cohorts.

Research Question 3: What factors are associated with students' risk to fail a course?

Regardless of the many possible definitions of student academic success, in this study, we focus on students' first module exam score for the following reasons. First of all, one of the purposes of the study is to identify the students at risk at an early stage, so that the study can help the course instructor set up an early warning system. It is necessary to use students' learning outcome at earlier stage rather than the final exam in order to advance the target student success variable. Secondly, according to one of the previous pilot studies, students' first exam score is one of the significant predictors of passing the course.

Instead of using the score of 60 as the cut-off for identifying students at risk, we used the score of 83, which is the lower quartile, as the cut-off score. There were 18.3% of students who either dropped out or failed the course according to the history data.

Table 7

Group Comparison on Pre-college Characteristics and Homework Learning Activity Features

	Range	Total		Gender					First Generation				
				Male (N=211)		Female (N=79)		t	Yes (N=44)		No (N=247)		t
		M	SD	M	SD	M	SD		M	SD	M	SD	
Studyskills	0-5	3.4	0.54	3.4	0.54	3.4	0.54	.96	3.3	0.43	3.4	0.56	-.44
M_financial	0-5	4.0	0.61	4	0.61	4.1	0.61	1.51	4.0	0.75	4.0	0.58	-.06
M_intrinsic	0-5	4.2	0.55	4.2	0.57	4.2	0.49	-.83	4.2	0.52	4.2	0.55	.01
M_societal	0-5	4.5	0.48	4.5	0.5	4.5	0.43	.45	4.5	0.38	4.5	0.50	.56
Self_efficacy	0-5	3.8	0.72	3.9	0.75	3.7	0.64	-1.81	3.7	0.69	3.9	0.73	-1.30
Eknowledge	0-3	1.40	0.36	1.4	0.36	1.4	0.36	-.92	1.3	0.25	1.4	0.37	-3.33**
Pearly	0-100	73.8	29.34	71.8	30.76	79.1	24.71	2.09*	73.2	30.39	73.9	29.21	-.15
Pregular	0-100	21.0	23.01	22.1	23.99	17.9	20.15	-1.50	20.2	24.42	21.1	22.80	-.25
Plate	0-100	5.2	12.45	6.1	13.89	3.0	7.05	-2.49*	6.6	15.71	4.9	11.80	.82
PansweredQ	0-100	97.5	9.36	96.8	10.73	99.3	3.38	3.02**	97.5	9.82	97.5	9.30	.03
Mattempts	0-3.6	1.7	0.36	1.7	0.38	1.8	0.31	2.01*	1.8	0.45	1.7	0.34	1.90
Correctness	0-100	96.4	10.15	95.6	11.5	98.4	4.55	3.04**	96.4	11.04	96.4	10.01	.00
Prepareness	0-100	61.0	11.73	60.8	12.42	61.5	9.81	.42	59.1	12.40	61.4	11.60	-1.18
LearningSpeed	0-3.3	1.7	0.31	1.7	0.32	1.8	0.29	1.32	1.8	0.37	1.7	0.30	1.89

Note. * $p<.05$, ** $p<.01$

Therefore, we believed that the lower quartile could be considered as a reasonable cut-off of students at risk. In order to identify possible predictors at early stage of the course, we used C5.0 decision tree (Kuhn & Johnson, 2013) to run classification analysis, using all of the pre-college characteristics (including the survey constructs and demographic variables) and online homework learning activity features as input variables.

Since the number of students not at risk is almost three times that of the number of students at risk, we first duplicated samples of students at risk three times to create a balanced dataset. This resulted in 411 samples in the root node of the tree. Then, the data were divided into two sets. 75% of the samples were used for training, and the other 25% were used for validation.

The decision tree obtained is shown in Figure 6. Students at risk are marked with red rectangle and value of 1. The misclassification rate is 22.38%. Among all the attributes, Correctness, Preparedness, Self-efficacy, and Plate were identified as the important predictors.

Research Question 4: What are the characteristics of students at risk?

The largest leaf node is Node 9, which has 97 samples classified as "students not at risk". The misclassification rate is 7.2%. Students in this group are featured with higher correctness rate, higher preparedness score, and never attempt to do the homework after deadline.

The next largest leaf node is Node 1, which has 75 samples classified as "students at risk." The misclassification rate is 17.3%. As we examine the tree structure, we can see that students that reach this node have low percentage of correctly answered questions

(Correctness<= 94.8%).

Other student-at-risk nodes are node 5 and node 12. Students in node 5 are featured with lower preparedness score, lower self-efficacy in math and science, and lower correct rate (94.8<= correctness < 99.3%). Students in node 12 are featured with very high correctness score (>99.3%) and higher preparedness score, but they attempted to do their homework after deadline. Table 8 summarized all of the leaf nodes with their sample size, misclassification rate and characteristics.

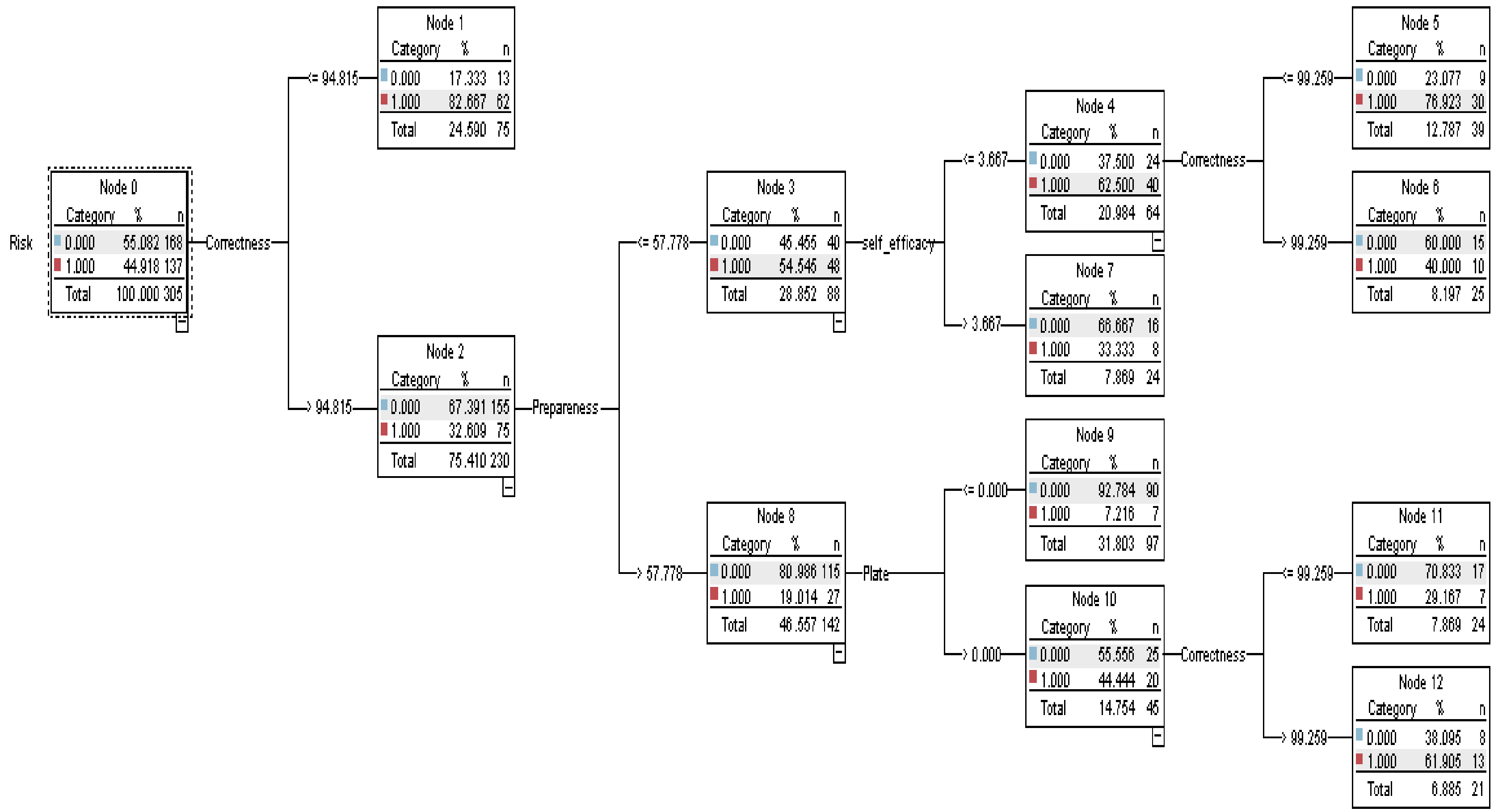

Figure 6. Decision Tree Model

Table 8
Summary of Leaf Nodes

ID	Sample Size	Characteristics	Classification	Misclassification Rate
1	75	Correctness <= 94.8%	Risk	17.3%
7	24	Correctness > 94.8% Preparedness<=57.8% Self-efficacy > 3.7	Non-risk	33.3%
9	97	Correctness > 94.8% Preparedness>57.8% Plate =0	Non-risk	7.2%
5	39	94.8%<=Correctness < 99.3% Preparedness<=57.8% Self-efficacy <=3.7	Risk	23.1%
6	25	Correctness > 99.3% Preparedness<=57.8% Self-efficacy <=3.7	Non-Risk	40.0%
11	24	94.8%<=Correctness < 99.3% Preparedness>57.8% Plate > 0	Non-Risk	29.2%
12	21	Correctness > 99.3% Preparedness>57.8% Plate > 0	Risk	38.1%

Chapter 5

Discussion

The purpose of this study was to create a working model to predict students at risk in an introduction to engineering course by considering students' pre-college characteristics, psychological traits, and online homework learning behavior features. It was hypothesized that the Engineering Students' Pre-college Characteristics survey was reliable and valid. Four research questions were examined in this study.

1. How actively are first year engineering students engaged in their online homework study?
2. Are there any group differences (gender and first generation) on students' pre-college characteristics, psychological traits, and online homework activities?
3. What factors are associated with students' risk to fail a course?
4. What are the characteristics of students at risk?

Validity and Reliability of the Engineering Students' Pre-college Characteristics Survey

The findings of this study support the hypothesis that the Engineering Students' Pre-college Characteristics survey is reliable and valid. Internal consistency reliability and Construct validity analyses were conducted to assess the reliability and validity of the survey instrument.

Confirmatory factor analysis (CFA) was used to examine the construct validity. As summarized before in the result section, fitting indices such as chi-square, chi-square to degree of freedom, goodness of fit index (GFI), Comparative fit index (CFI), and Root mean square error of approximation (RMSEA) were reported to confirm the structure of the instrument constructs. In addition to the fitting index, significant factor loadings and insignificant covariate among constructs were also considered as supporting evidence of construct validity.

When first examining the hypothesized structure of the survey constructs, although the chi-square to degree of freedom and RMSEA met the cut-off criteria stated by Hu and Bentler (1999), GFI and CFI value did not meet the 0.9 criteria. As indicated by the regression weights, there were two constructs that were not stable and valid: mentor motivation and parental motivation. Item loadings on those two factors were not significant. Then the author ran a new model without those two constructs. The factor loadings of the new model were all significant. Furthermore, all of the model fit index met the cut-off criteria claimed by previous research studies (Hu & Bentler, 1999; Kline, 2010).

The study also summarized the internal consistent reliability by calculating the Cronbach's alpha values. Previous literature suggested that the value of a Cronbach's Alpha over .70 is acceptable internal consistent reliability in social science studies (Bland & Altman, 1997; Tavakol & Dennick, 2011; George & Mallery, 2003). George and Mallery (2003) provide more specific rules of thumb when explaining Cronbach's Alpha:

- Alpha > .9 – Excellent,
- Alpha > .8 – Good,
- Alpha > .7 – Acceptable
- Alpha > .6 – Questionable,
- > .5 – Poor, and_
- Alpha < .5 – Unacceptable".

The majority of the constructs in the modified model have internal reliability larger than 0.70 except societal motivation (Cronbach's Alpha = .69) and engineering knowledge (EKnowledge) (Cronbach's Alpha = .57). Although those two constructs have questionable or poor internal reliability, the author decided to keep the two constructs in the follow-up analysis for the following reasons. Firstly, the societal constructs had acceptable internal reliability in larger national studies (Haase et al., 2013; Sheppard et al., 2010). For example, Haase et al. (2013) applied the survey in a cross-national study to measure US and Denmark engineering student profiles. The internal reliability of those societal motivation constructs were over .70 in both US and Denmark samples. Secondly, some researchers suggest 0.60 as the cut off of acceptable reliability in social science studies (e.g., Sheppard et. al., 2010). Thus we could accept the societal motivation construct as a reliable construct). 3. When Sheppard el al.(2010) developing the APPLES survey instrument, they considered the exposure to engineering professionals/education construct as a single item and kept it in their instrument.

In summary, the majority of the constructs in the survey instrument used in the present study have acceptable construct validity and internal reliability. The two constructs, namely parental motivation and mentor motivation, did not demonstrate acceptable construct validity, thus they were excluded from the follow-up analysis.

Research Question 1: How actively are first year engineering students engaged in their online homework study?

The research findings indicated that students in EF151 Introduction to Physics were actively engaged in their homework study in different ways: first of all, students made use of the early completion bonus that they had majority of their homework attempts one day earlier than the deadline. Secondly, students tried their best to answer all of the homework questions that the online homework system recorded an average of 97.5% questions attempted by students. Finally, students have more than one attempt at each homework questions averagely.

According to the literature, engagement in homework is the key part of students' out-of-class study experience especially in STEM area, where large quantity of computation and practice are required (Palocsay & Stevens, 2008; Cooper, Robinson, & Patall, 2006). Specifically, the EF151 course instructor found that students score 80 or above on the homework had 95.9% passing rate on the course while students score below 80 on the homework only had 33.3% passing rate on the course (Bennett et al., 2013). Therefore, it is essential to know how students engaged in the online homework study to develop a working model to identify students at risk. The most common ways to investigate students' engagement in homework are self-reported study hours and homework performance measures (Cheng, Thacker, Cardenas, & Crunch, 2004; Peng, 2009). Although the online homework system keeps track of the students' learning activities, the large quantity of data is rarely utilized to describe students' engagement on homework.

The present study applied data mining technique to extract students' homework learning activity features such as learning speed (LearningSpeed) and homework attempts one day earlier than the deadline (Pearly). The extracted feature enabled the researchers to describe students' engagement on homework in different ways. As indicated by the findings, students were actively engaged in the homework studies that they started to work on their homework early, had multiple attempt to get to the right answers, and tried their best to answer the homework questions.

Research Question 2: Are there any group differences (gender and first generation) on students' pre-college characteristics, psychological traits, and online homework activities?

The study found no significant differences on male and female students' motivation to study in Engineering, self-rated study skills, self-efficacy in math and science, and pre-college exposure to engineering professionals. This result is not consistent with the literature where gender differences were commonly found in those constructs. For example, Vogt, Hocevar, and Hagedorn (2007) conducted a study to investigate the gender difference on engineering students' self-confidence and self-efficacy and found that males had higher scores on both of the two constructs. The inconsistency of the study results might because the small female sample included in the present study was not representative of all of the female engineering students. Another possible reason was that along with the development o the society and emphasis in female success in the area of engineering, female stuents are more actively engaged in their engieering study and more confident in their personal engineering competences. The female students in Fall 2015 EF151 course (our sample) are equal to their male peers on those pre-college characteristics. However, gender differences were found on the online homework learning activity variables that female students were more actively engaged.

First generation engineering students were frequently reported to be different from traditional engineering students in multiple academic and social integration processes (Fernandez, Trenor, Zerda, & Cortes, 2008; Collier & Morgan, 2008; Engle & Tinto, 2008). The group differences found by Engle and Tinto (2008) were intrapersonal communication with other students and faculty collaborate learning, participating in extracurricular activities, and using support services. Those were not the aspects we examined in the current study. The results of the current study, although not supportive of the stereotype of first generation students' lower performance at college than other students, illustrated that first generation students, who worked as proactively as their cohorts, had the same performance level. The only difference found in the current study was level of exposure to engineering professionals. One of the survey items in this construct asked students whether or not their parent/siblings hold an engineering degree. Students who are first generation in their family to go to college should not have a yes answer to this question, which definitely would influence their score on this construct. In addition, first generation students stated that they have lower levels of knowledge about engineering before college.

Research Question 3: What factors are associated with students' risk to fail a course?

Using the lower quartile of exam 1 score as the cut-off of students-at-risk and other students, we built a decision tree model (Kuhn & Johnson, 2013) considering all of the demographic variables, survey constructs, and online homework activity features. The final model selected Correctness, Preparedness, Self-efficacy, and Plate as the important predictors. The misclassification rate of the model is 22.38%.

The decision tree model confirmed Veenstra et al.'s proposed first year engineering retention model that both students' pre-college characteristics (self-efficacy) and students' college academic integration process (Correctness, Preparedness, and Plate) influenced students' learning outcome (Veenstra et al., 2009). Self-efficacy in math and science has been suggested by many empirical studies to be the important predictor of engineering students learning outcome (Veenstra et al., 2008; Veenstra et al., 2009). As quantitative computation and scientific thinking are two major components in most of the engineering core courses, students' self-efficacy in math and science impacted their performance in the courses. As summarized in Chapter 1, although Veentra et al. (2009) believed that college integration process was influential to engineering students' learning outcome, they did not incorporate any variables in their empirical studies to support the proposed model. Even though we did not examine the whole college integration process, we applied a data mining approach to observe engineering students' online homework learning process and found the association between students' homework learning and academic outcome. The results were consistent in Minaei-Bidgoli et al.'s data mining study (Minaei-Bidgoli, Kashy, Kortemeyer, & Punch, 2003).

Minaei-Bidgoli et al. extracted similar features such as correctness, average attempts on homework problems, and preparedness to predict students' final exam scores in an Introductory Physics Course for Scientists and Engineers. They found that Total_Correct_Answer, Total_Number of_Tries, First_correct_Answer, Time_Spent_to_Solve, Total_Time_Spent and Communication were indicators of students' passing rate. Similarly we both found students' correct answer was the most important factor and Preparedness (First_Correct_Answer) was another important factor. We did not extract the feature of working hours (Time_Spent_to_Solve and Total_Time_Spent) in our study because the online homework system did not track students' log out timestamp. We were unable to keep track of students' working hours on the homework problems. Instead we extracted Pearly, Pregular, and Plate as the indicator of students' proactiveness and procrastination on homework learning process. Interestingly we've found whether or not students procrastinated on homework (Plate >0 or Plate =0) was one of the factors impacting students' learning outcome. Students in the non-risk group do not have homework completed after deadline (Plate = 0). There was another inconsistency between our study result and Minaei-Bidgoli et al.'s results (Minaei-Bidgoli et al., 2003). The total/average attempts variable was not influential in our study while it was the second important factor in Minaei-Bidgo et al.'s study. This inconsistency in the result brought us to the discussion of the decision tree variable selection mechanism.

From an algorithmic point of view, building a decision tree has two major procedures. Firstly, it includes a forward stepwise procedure that adds feature variables. Then the second procedure comes to a pruning process (Moon et al., 2012). The severity of pruning process significantly influenced the variables included in the final model and the misclassification rate. The software used in this model, SPSS Modeler 15.0, has a default setting for a decision tree. The pruning criteria of the default setting allows 2 and above cases in each single node. However, this pruning severity would result in a very complicated tree, which would have a tremendous number of leaf nodes. Most of the time the model ran under the default setting will have a relative low misclassification rate (<10%). However, the complexity of the model makes it hard to explain. Furthermore, from a cost benefit point of view, identifying 2 students in an at-risk group and developing targeted intervention for only 2 students will be time-consuming and costly. Therefore, we change the default setting to prune at least 15 students in each node, which is about 5% of the sample size. The difference pruning severity between our study and

Minaei-Bidgo et al.'s study might be the possible reason why we have inconsistent result. In fact, when we ran the model using the default setting of the software, we found the first generation and father's education were also important factors.

In summary, the present study built a decision tree model to support Veenstra et al's proposed first year engineering retention model (Veenstra et al., 2009). In order to improve the explanability of the decision tree model, the pruning criteria was set up as more than 15 students in each leaf node. The result of the decision tree identified 4 factors associated with students-at-risk in the course. Those 4 factors were: Correctness, Preparedness, Self-efficacy, and Plate. Students need to be actively engaged in their homework study, for example, complete their homework earlier rather than procrastinated until after deadline, to succeed in the course. Furthermore, students should better prepare themselves for the homework and get higher correct rate on homework through pre-lecture learning and in-class learning. The results supported the hypothesized model that both students' pre-college characteristics and college integration were influential to students' academic success.

Research Question 4: What are the characteristics of students at risk?

There were three types of students-at-risk identified by the decision tree model. The three groups of students nested in node 1, 5 and 12 respectively. Students in node 1 were featured with very low correct rate on homework questions. Possible intervention for this group of students might be one-on-one tutor and additional help session customized for homework problems. Students in node 5 had a medium correct rate, low preparedness score, and low self-efficacy. They could not get the first attempt at homework question correct but still achieved a medium correction rate. Students in this group were actively engaged in homework study in that they attempted to solve the problem multiple times in order to increase their correct rate, but they've experienced problems and struggled in the process. Collaborative learning or other effective learning strategies may be encouraged for this group of students to facilitate their learning process in a more efficient way. In addition, student success center may provide supplemental instruction to this group of students. Students in node 12 were featured with very high correct rate and high preparedness score, but they procrastinated on homework that they attempt to solve until after deadline. Course instructors may send email notifications to encourage this group of students to actively engage in homework and complete their homework earlier.

When we take a look at the students who are not at risk, we can also identify their characteristics from the decision tree model: although varied by Preparedness score and self-efficacy, they have relative high correct rate at homework and only a very small portion of the non-risk students (24 out of 170) worked on their homework after deadline. It is necessary to encourage students complete their homework in a more proactive way – making use of homework bonus and finishing homework earlier. The results of the study are good illustrations to show students the importance of proactiveness on homework.

Practical Implication

The study provided practical implication for the EF 151 Introduction to Physics course at the University of Tennessee, and the overall engineering education in several aspects. First of all, the study validated an instrument to investigate engineering students' pre-college characteristics. In addition to college entrance screening criteria such as ACT/SAT scores, faculty and staff could have better understanding of their students in various aspects such as exposure to engineering professionals before college, motivation to study in engineering, and self-efficacy in math and science.

Secondly, as higher institutions are all approaching the incorporation of online course management systems and online homework, large quantities of data have been recorded in the course management system. However, without applying data mining techniques, course instructors often leave the quantity of information unanalyzed. Course instructors only used the course system to keep track of students' performance system and even though some course established student profiles within the course system, the profile summaries rarely made use of students' activity log data. Those student profiles only contain student background information and performance measure records. The study applied data mining technique to help the course (EF151 and other Engineering course made use of CMS) to establish students' profile in an innovative way that applied statistics algorithm to transform students' activity log data into indicators of students' engagement in learning activities. In this way, course instructors, teaching assistant, and students themselves can better observe how students engaged in the learning process and how they can improve their learning. For example, students may have more attempts on homework problem or attempt to working on homework problems earlier.

Thirdly, the study developed a working model to identify students-at-risk at the early stage of the course. The model can be applied in the course to build a warning system. Course instructors can identify those students at early stage and develop corresponding intervention to help the students. EF151 course instructors used to send emails to notify the students whose average homework scores were under 80. The content of the email was reminding students of the available help resources such as student help session and Supplemental Instruction. However, only homework performance was the judging criteria and no targeted interventions were developed. The model developed in the present study can help build a more reliable warning system taking consideration of different influential factors and group students-at-risk with similar characteristics. As a result, customized email notification and other intervention techniques can develop to help students-at-risk in different groups.

Finally, combined with the course warning system, course instructors can make use of the model to develop targeted intervention to help students-at-risk. There are tremendous resources students can choose to help their homework learning in EF151. For example, the course provided old exams from previous years for students to practice; the course provide student help session in different location to help students with any questions about the course content and homework; lecture notes and online tutorials are available in the course website; discussion boards are available in the course website for students to discuss questions and problems with the instructor and TAs, etc. One student cannot make use of all of the resources and it is essential for both the course instructors and students to understand students' real needs and choose the right resources. As discussed in previous sections, possible customized intervention can be developed to help students with different characteristics. In this way, students can improve their homework learning in a more efficient way. Similar approaches can also be applied to other engineering courses with online course management settings.

In summary, the present study identified reliable and valid instrument and approaches to help engineering course instructors to build up student profiles identify students-at-risk at earlier stage by building up statistical models, and the result of the model can facilitate engineering educators to customize intervention with different type of students-at-risk. Engineering students can have a better understanding of their learning status and improve their learning in more efficient ways.

Limitations of the Present Study

There are a number of limitations to this current study. First of all, although the present study collected information about students' pre-college characteristics and engineering homework learning behaviors, the variables are not inclusive of all related feature variables. Starting from Tinto's interactionalist theory model (Tinto, 1993) and Veenstra et al.'s first year engineering retention model (Veenstra et al., 2009), social integration is one of the most important aspect of students' college integration process and influential of students' learning outcome. The social integration process relates to how students interact with faculty and peers at college, participate in extracurricular activities, and deal with interpersonal relationships. The current study only investigated a very small portion of engineering students' college integration process-online homework learning process in an Engineering course while left students' other learning process such as in-class learning, and social integration variables.

Secondly, because of the limitation of sample size and pruning mechanism of decision tree model, the current model is limited to certain variables, which made the statistical model not inclusive and reliable enough. When we changed the pruning settings into the default software setting, which allows down to 2 participants in a node, first generation and father's education level became two important factors. If we have enough sample size, we might be able to accurately identify the relationships between the predictors and students' learning outcome while still keep the decision tree model explainable and practical.

Thirdly, the participant recruitment process and sampling procedure may result in some biases. For example, students who had lower performance or are at risk to fail the course may be less willing to release their online homework log to the researcher. As a result, the study has the potential risk to identify certain at risk students.

Future Research

Future studies should be able to address the limitation of the current study. First of all, a more comprehensive survey instrument should be developed and validated to examine both students' pre-college characteristics and students' social integration process. Personal response system (clicker system) can be used to record students' engagement in lectures and the recorded data can contribute to the description of student profiles. Furthermore, it would be helpful for the researchers to analyze students' communication with other students, TAs, and course instructors in the online discussion board applying text-mining techniques. After collecting more comprehensive information about students, larger sample size should be recruited to improve the accuracy and reliability of the statistic model. Cumulative data from multiple enrolled cohorts can help enlarge the sample size if the course setting did not change dramatically.

When a reliable and accurate working model is developed to predict the students at risk and corresponding intervention strategies are developed to help the those students. It is necessary to conduct an experimental design research to examine the targeted intervention approach. In one way, the empirical study can support and validate the developed statistical model. In the other way, experimental design can help build up the causal relationship between the targeted intervention and students' learning outcome.

Students in EF151 course can be randomly divided into treatment group and control group. Students in treatment group will receive customized intervention while students in control group will receive regular email notifications to remind them of the available resources. Comparisons of those two groups on final exams will be conducted to examine the effectiveness of customized intervention.

Conclusion

In the present study, a reliable and valid instrument to measure engineering students' pre-college characteristics has been developed. The study also applied data mining approach to analyze the online homework logs in order to observe engineering students' homework learning process. A decision tree model containing all of the pre-college characteristics and online homework learning features has been developed to identify four key factors related to students' risk to fail the first model exam: Correctness, Preparedness, Self-efficacy, and Plate. The result of the decision tree model helped identify students-at-risk at early stage of the course. Students-at-risk were grouped in to different groups. The author also proposed customized intervention to help students in different groups. The findings of the study helped engineering students and educators to build up comprehensive student profile and better understand students' status and learning needs. Thus both the engineering educators and students can help improve the learning process more efficiently.

List of References

Allison, P. D. (2001). *Missing Data.* Sage University Papers Series on Quantitative Applications in the Social Sciences. 07-136. Thousand Oaks, CA: Sage.

Ahmad, A., Apandi, N. I. A., Hashim, M. R., Sani, Z. M., Anuar, H., & Khalil, S. N. (2012). A study on students' interest and preparedness based on entry survey. In *2012 International Conference on Statistics in Science, Business, and Engineering (ICSSBE)* (pp. 1–4). doi:10.1109/ICSSBE.2012.6396530

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of college student personnel*, *25*(4), 297-308.

Astin, A. W., & Oseguera, L. (2005). Degree attainment rates at American colleges and universities: Revised edition. *Higher Education Research Institution, University of California Los Angeles*.

Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ*, *1986*.

Bennett, R., Li, W., Olsen, T., Schleter, W., & Guffey, S. (2013 June). *Characteristics of Students Who Do Not Do Homework.* Poster at American Society of Engineering Education (ASEE), Atlanta, GA.

Bennett, R. M., Schleter, W., & Raman, D. R. (2012). *A Success Enhancement Program after the First Test in Freshman Engineering.* Poster at American Society of Engineering Education (ASEE), San Antonio, TX.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588.

Bernold, L. E. (2007). Preparedness of engineering freshman to inquiry-based learning. *Journal of Professional Issues in Engineering Education and Practice*, *133*(2), 99-106.

BlackBoard (2007). <http://www.blackboard.com/>.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj*, *314*(7080), 572.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*(3), 303-316.

Boyd, C. J., Frey, M. A., & Aaronson, L. S. (1988). Structural equation models and nursing research: Part I. *Nursing research*, *37*(4), 249-251.

Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology*, *93*(1), 55–64. http://doi.org/10.1037/0022-0663.93.1.55

Chen, G. D., Liu, C. C., Ou, K. L., & Liu, B. J. (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research*, *23*(3), 305-332.

Cheng, K. K., Thacker, B. A., Cardenas, R. L., & Crouch, C. (2004). Using an online homework system enhances students' learning of physics concepts in an introductory physics course. *American Journal of Physics*, *72*(11), 1447–1453. http://doi.org/10.1119/1.1768555

Clough, G. W. (2004). The engineer of 2020: Visions of engineering in the new century. *National Academy of Engineering, Washington, DC*.

Collier, P. J., & Morgan, D. L. (2008). “Is that paper really due today?”: differences in first-generation and traditional college students’ understandings of faculty expectations. *Higher Education*, *55*(4), 425-446.

Colton, D., & Covert, R. W. (2007). *Designing and constructing instruments for social research and evaluation*. John Wiley & Sons.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of educational research*, *76*(1), 1-62.

Dimitrov, D. M. (2008). *Quantitative research in education*. New York, NY: Whittier Publications Inc.

Eagan, K., Hurtado, S., & Chang, M. (2010, October). What matters in STEM: Institutional contexts that influence STEM bachelor’s degree completion rates. In *annual meeting of the Association for the Study of Higher Education, Indianapolis, IN*.

Eccles, J. S., & Wigfield, A. (1995). In the Mind of the Actor: The Structure of Adolescents’ Achievement Task Values and Expectancy-Related Beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215–225.

Eccles, J., T.F. Adler, R. Futterman, S.B. Goff, C.M. Kaczala, J.L. Meece, and C. Midgley. (1983). *Expectancies, values, and academic behaviors.* In *Achievement and achievement motivation. ed*. J. T. Spence, 75–146. San Francisco, CA: Freeman.

Engle, J., & Tinto, V. (2008). Moving Beyond Access: College Success for Low-Income, First-Generation Students. *Pell Institute for the Study of Opportunity in Higher Education*.

Fernandez, M. J., Trenor, J. M., Zerda, K. S., & Cortes, C. (2008, October). First generation college students in engineering: A qualitative investigation of barriers to academic plans. In *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual* (pp. T4D-1). IEEE.

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.

Furr, R.M., & Bacharach, V.R. (2014). *Psychometrics: An Introduction*. (2nd ed.). Los Angeles, CA: Sage Publications.

Gaudioso, E., & Talavera, L. (2006). Data mining to support tutoring in virtual learning communities: Experiences and challenges. (pp. 207–226). Wit Press.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Goulet, R. (2010). Individualized homework: an effective learning strategy. In *American Society for Engineering Education*. American Society for Engineering Education.

Haase, S., Chen, H. L., Sheppard, S., Kolmos, A., & Mejlgaard, N. (2013). What Does It Take to Become a Good Engineer? Identifying Cross-National Engineering Student

Profiles According to Perceived Importance of Skills. *International Journal of Engineering Education*.

Hirsch, L. S., Kimmel, H., Rockland, R., & Bloom, J. (2005). Implementing pre-engineering curricula in high school science and mathematics. In *Frontiers in Education, 2005. FIE '05. Proceedings 35th Annual Conference* (p. S2F–21).

Hirsch, L. S., Kimmel, H., Rockland, R., & Bloom, J. (2006). Using Pre-engineering Curricula in High School Science and Mathematics: A Follow-up Study. In *Frontiers in Education Conference, 36th Annual* (pp. 1–6).

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Huo, X., Kim, S. B., Tsui, K. L., & Wang, S. (2006). FBP: a frontier-based tree-pruning algorithm. *INFORMS Journal on Computing*, *18*(4), 494-505.

Jones, B. D., Paretti, M. C., Hein, S. F., & Knott, T. W. (2010). An Analysis of Motivation Constructs with First-Year Engineering Students: Relationships Among Expectancies, Values, Achievement, and Career Plans. *Journal of Engineering Education*, *99*(4), 319–336.

Keup, J.R. (2004). The cooperative institutional research program freshman survey and your first college year: using longitudinal data to assess the first year of college. *Assessment Update, 16*(2), 8-10.

Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage Publications.

Knight, M., & Cunningham, C. (2004, June). Draw an engineer test (DAET): Development of a tool to investigate students’ ideas about engineers and engineering. In *ASEE Annual Conference and Exposition* (Vol. 2004).

Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003, January). Preventing student dropout in distance learning using machine learning techniques. In *Knowledge-Based Intelligent Information and Engineering Systems* (pp. 267-274). Springer Berlin Heidelberg.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (pp. 61-92). New York: Springer.

Lackey, L. W., Lackey, W. J., Grady, H. M., & Davis, M. T. (2003). Efficacy of Using a Single, Non-Technical Variable to Predict the Academic Success of Freshmen Engineering Students. *Journal of Engineering Education*, *92*(1), 41–48.

LaNasa, S. M., Cabrera, A. F., & Trangsrud, H. (2009). The construct validity of student engagement: A confirmatory factor analysis approach. *Research in Higher Education*, *50*(4), 315-332.

Lent, R. W., Sheu, H.-B., Singley, D., Schmidt, J. A., Schmidt, L. C., & Gloster, C. S. (2008). Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *Journal of Vocational Behavior*, *73*(2), 328–335. http://doi.org/10.1016/j.jvb.2008.07.005

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.

Liberatore, M. W. (2011). Improved Student Achievement Using Personalized Online Homework. *Chemical Engineering Education*, *45*(3), 184-190.

Lotkowski, V. A., Robbins, S. B., & Noeth, R. J. (2004). *The Role of Academic and Non-Academic Factors in Improving College Retention. ACT Policy Report*. Retrieved from http://eric.ed.gov/?id=ED485476

Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003). Predicting student performance: an application of data mining methods with an educational Web-based system. In *Frontiers in Education, 2003. FIE 2003 33rd Annual* (Vol. 1, p. T2A–13). http://doi.org/10.1109/FIE.2003.1263284

Moodle (2007). <http://moodle.org/>.

Moon, S. S., Kang, S. Y., Jitpitaklert, W., & Kim, S. B. (2012). Decision tree models for characterizing smoking patterns of older adults. *Expert Systems with Applications*, *39*(1), 445-451.

Morrow J. A. & Skolits G. (2014 October). *Twelve Steps of Data Cleaning: Strategies for Dealing with Dirty Evaluation Data.* Workshop at American Evaluation Association (AEA), Denver, CO.

Nowicki, S., & Duke, M. (1992). The association of children's nonverbal decoding abilities with their popularity, locus of control, and academic achievement. *Journal of Genetic Psychology*, *153*, 385–393.

Ohland, M.W., S.D. Sheppard, G. Lichtenstein, O. Eris, D. Chachra, & R.A. Layton. (2008). Persistence, engagement, and migration in engineering program. *Journal of Engineering Education, 97* (3): 259–78.

Palocsay, S. W., & Stevens, S. P. (2008). A Study of the Effectiveness of Web-Based Homework in Teaching Undergraduate Business Statistics. *Decision Sciences Journal of Innovative Education*, *6*(2), 213-232.

Peng, J. C. (2009). Using an Online Homework System to Submit Accounting Homework: Role of Cognitive Need, Computer Efficacy, and Perception. *Journal of Education for Business*, *84*(5), 263–268. http://doi.org/10.3200/JOEB.84.5.263-268

Phase, I. I. (2005). *Educating the Engineer of 2020: Adapting Engineering Education to the New Century*. National Academies Press.

Raykov, T., & Marcoulides, G. A. (2006). Estimation of generalizability coefficients via a structural equation modeling approach to scale reliability evaluation. *International Journal of Testing*, *6*(1), 81-95.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. http://doi.org/10.1016/j.compedu.2007.05.016

Rugarcia, A., Felder, R. M., Woods, D. R., & Stice, J. E. (2000). The future of engineering education I. A vision for a new century. *Chemical Engineering Education*, *34*(1), 16-25.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Statistical conclusion validity and internal validity. *Experimental and quasi-experimental designs for generalized causal inference*, 45-48.

Sheppard, S., Gilmartin, S., Chen, H. L., Donaldson, K., Lichtenstein, G., Eris, O., Lande, M., & Toye, G. (2010). Exploring the Engineering Student Experience:

Findings from the Academic Pathways of People Learning Engineering Survey (APPLES). TR-10-01. *Center for the Advancement of Engineering Education (NJ1)*.

Schilling, W. (2010). Using performance bonuses to decrease procrastination. In *American Society for Engineering Education*. American Society for Engineering Education.

Schleter, W. & Bennett, R. (2006). Using Web-based Homework in an Introductory Engineering Physics Course. In *American Society for Engineering Education*. American Society for Engineering Education.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston, MA: Allyn and Bacon.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53.

Tella, A., & Tella, A. (2011). Reliability and Factor Analysis of a Blackboard Course Management System Success: A Scale Development and Validation in an Educational Context. *Journal of Information Technology Education: Research*, *10*(1), 55–80.

Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher Education*, *37*(1), 1–22.

Tinto, V. (1993). Leaving college rethinking the causes and cures of student attrition.: The University of Chicago. *Chicago, IL.*

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.

Veenstra, C. P., Dey, E. L., & Herrin, G. D. (2009). A model for freshman engineering retention. *Advances in Engineering Education*, *1*(3), 1-31.

Vogt, C. M., Hocevar, D., & Hagedorn, L. S. (2007). A social cognitive construct validation: Determining women's and men's success in engineering programs. *The Journal of Higher Education*, *78*(3), 337-364.

Weiss, T. (2009). The 10 hardest jobs to fill in America. Retrieved from http://www.forbes.com/2009/06/03/hard-jobs-fill-leadership-careers-employment.html.

Wohlgemuth, D., Whalen, D., Sullivan, J., Nading, C., Shelley, M., & Wang, Y. (Rebecca). (2007). Financial, Academic, and Environmental Influences on the Retention and Graduation of Students. *Journal of College Student Retention: Research, Theory and Practice*, *8*(4), 457–475.

Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher Education*, *37*(1), 1–22. http://doi.org/10.1007/BF01680039

Zajacova, A., Lynch, S. M., & Espenshade, T. J. (2005). Self-Efficacy, Stress, and Academic Success in College. *Research in Higher Education*, *46*(6), 677–706. http://doi.org/10.1007/s11162-004-4139-z

Zorrilla, M. E., Menasalvas, E., Marín, D., Mora, E., & Segovia, J. (2005). Web Usage Mining Project for Improving Web-Based Learning Sites. In R. M. Díaz, F. Pichler, & A. Q. Arencibia (Eds.), *Computer Aided Systems Theory – EUROCAST 2005* (pp. 205–210). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/11556985_26

Appendices

Appendix A

Data Request Description

Take all students enrolled in EF151 in Fall 2015, prepare the following tables:

1. Student outcome: exam grades

StudentID	Exam1

- Each record corresponds to one student. So StudentID is the key for this table, no duplicate should exist.

2. Student activity

Each attempt of a homework question:

StudentID	HomeworkQuestionID	TimeStamp	Result

- HomeworkQuestionID: specific to subquestions that are gradable, e.g., hw 1-8-3, hw 1-8-4, etc.
- TimeStamp: the time of this student's attempt. Each subquestion may have multiple records, each with a different time stamp.
- Result: binary -- right or wrong for that attempt.

3. Homework Information

Each Homeworkset:

HomeworkSetID	Assign Date	Due Date

Appendix B

Engineering Students' Pre-college Characteristics Survey

INTRODUCTION

You are invited to participate in a research study at the University of Tennessee. The objective of the research is to develop an engineering student success model to identify factors associated with engineering students' success and use the model to develop possible interventions to help students who are at risk of failing the course.

It will take your 15-20 minutes to complete the survey and you **will get extra credit for completing the survey.** You will be requested to enter your name in the survey as you consent to participate. The principal investigator (Wenshu Li) will provide your names to your instructor to process the extra credits. If you decide not to participate in the survey, you won't have any grades deduction and you can still complete your homework earlier to earn extra credits.

INFORMATION ABOUT PARTICIPANTS INVOLVEMENT IN THE STUDY

The survey will take you 15-20 minutes to complete, and will request information about your high school and college experience, in particular studying in the College of Engineering at the University of Tennessee, and your demographic information.

RISKS

The risks are minimal to the participants. Participants may take the risk of breach of confidentiality, but the researcher will protect against the risk. The procedures to protect participants' confidentiality are described in the Confidentiality section.

BENEFITS

There is no direct benefit for the participants. The result of the study will help the engineering educator to develop targeted interventions to help first year engineering students to improve their academic success. Furthermore, the result of the study will provide implications to general engineering curriculum development and modification to better cultivate future engineers.

CONFIDENTIALITY

All of the information will be kept confidential. Only the researchers have access to the information you enter. The data will be stored securely in a password required website. There will be no names and other identifiable information in the oral or written reports.

CONTACT INFORMATION

If you have questions at any time about the study or the procedures you may contact the researchers:

Wenshu Li

wli23@utk.edu

(951)801-8437

Dr. Richard Bennett

rmbennett@utk.edu

(865)974-7540

If you have questions about your rights as a participant, contact the Office of Research Compliance Officer at (865) 974-3466.

PARTICIPATION

Your participation in this study is voluntary. You may decline or withdraw to participate in the study at any time without penalty.

CONSENT

I have read and understand the above information. Please save a copy of this information for your records.

Please select your choice below.

Clicking on the "agree" button below indicates that:

• You have read the above information

• You voluntarily agree to participate

• You are at least 18 years of age

If you do not wish to participate in the research study, please decline participation by clicking on the "disagree" button.

- Agree
- Disagree

Your name will be used as a signature of your agreement to participate in this study. The research will use your name and your UTK email address to send you the instruction to earn extra credit and preserve your anonymity.

Your first name______________________

Your Last Name_____________________

Part I. Demographic Information. Please answer the following questions or circle the answer that best describe you and your family.

1. What is your age? ________

2. I am___

   a. Female

   b. Male

   c. Prefer not to answer

3. I am a/an ____(Select all that apply).

   a. American Indian

   b. Asian

   c. Pacific Islander

   d. African American

   e. Multiracial

   f. Caucasian

   g. Other (please specify) _______________

   h. Prefer not to answer

4.Are you Hispanic? ___

   a. Yes ___

   b. No

   c. Prefer not to answer

5. Are you the first generation in your family to attend college?

   a. Yes ___

   b. No

   c. Prefer not to answer

6. Are you enrolled as a full time or part-time student

   a. Yes ___

b. No

c. Prefer not to answer

7. What is the highest education level of your mother?

   a. Junior high/middle school or less

   b. Some high school

   c. High school graduate

   d. Postsecondary school other than college

   e. Some college

   f. College degree

   g. Some graduate school

   h. Graduate degree

8. What is the highest education level of your father?

   a. Junior high/middle school or less

   b. Some high school

   c. High school graduate

   d. Postsecondary school other than college

   e. Some college

   f. College degree

   g. Some graduate school

   h. Graduate degree

9. **What is your best estimate of your parents'/ guardians' total income last year? Consider income from all sources before taxes.** (Mark one)

   a. Less than $10,000

   b. $10,000-14,999

c. $15,000-19,999

d. $20,000-24,999

e. $25,000-29,999

f. $30,000-39,999

g. $40,000-49,999

h. $50,000-59,999

i. $60,000-74,999

j. $75,000-99,999

k. $100,000-149,999

l. $150,000-199,999

m. $200,000-249,999

n. $250,000 or more

Part II. We are interested in knowing your study habits. Please rate the following statements by circling the number. 0 stands for never and 4 stands for always.

	How often in the past year did you?	Never	Rarely	Sometimes	Often	Always
1	Ask questions in class	0	1	2	3	4
2	Support your opinions with a logical argument	0	1	2	3	4
3	Seek solutions to problems and explain them to others	0	1	2	3	4
4	Revise your papers to improve your writing	0	1	2	3	4
5	Evaluate the quality or reliability of information you received	0	1	2	3	4
6	Take a risk because you feel you have more to gain	0	1	2	3	4
7	Seek alternative solutions to a problem	0	1	2	3	4
8	Look up scientific research articles and resources	0	1	2	3	4
9	Explore topics on your own, even though	0	1	2	3	4

	it was not required for a class					
10	Accept mistakes as part of the learning process	0	1	2	3	4
11	Seek feedback on your academic work	0	1	2	3	4

Part III. We are interested in knowing why you are or were studying engineering. Please rate the following statements by circling the number. 1 stands for strongly disagree and 5 stands for strongly agree.

		Strongly Disagree	**Disagree**	**Neither Disagree Nor Agree**	**Agree**	**Strongly Agree**
1	Technology plays an important role in solving society's problems	1	2	3	4	5
2	Engineers make more money than most other professionals	1	2	3	4	5
3	My parent(s) would disapprove if I chose a major other than engineering	1	2	3	4	5
4	Engineers have contributed greatly to fixing problems in the world	1	2	3	4	5
5	Engineers are well paid	1	2	3	4	5
6	My parent(s) want me to be an engineer	1	2	3	4	5
7	An engineering degree will guarantee me a job when I graduate	1	2	3	4	5
8	A faculty member, academic advisor, teaching assistant or other university affiliated person has encouraged and/or inspired me to study engineering	1	2	3	4	5
9	A non-university affiliated mentor has encouraged and/or inspired me to study engineering	1	2	3	4	5
10	A mentor has introduced me to people and opportunities in engineering	1	2	3	4	5
11	I feel good when I am doing engineering	1	2	3	4	5
12	I like to build stuff	1	2	3	4	5
13	I think engineering is fun	1	2	3	4	5
14	Engineering skills can be used for the good of society	1	2	3	4	5

15	I think engineering is interesting	1	2	3	4	5
16	I like to figure out how things work	1	2	3	4	5

Part IV. Rate yourself on each of the following traits as compared to your classmates. We want the most accurate estimate of how you see yourself.

		Lowest 25%	Below Average	Average	Above Average	Highest 25%
1	Math ability	1	2	3	4	5
2	Science ability	1	2	3	4	5
3	Ability to apply math principles in solving real world problems	1	2	3	4	5
4	Ability to apply science principles in solving real world problems	1	2	3	4	5

Part IV. We are interested about your experience and knowledge about Engineering before college. Please answer the following question and choose the statement that best describe you experience.

1. Before college, how much knowledge did you have about the engineering profession?

a. No knowledge
b. Limited knowledge
c. Moderate knowledge
d. Extensive knowledge
e. I prefer not to answer

2. How much exposure have you had to a professional engineering environment as

How much exposure have you had to a professional engineering environment as	Not at All	Limited	Moderate	Extensive
A visitor	1	2	3	4
An intern	1	2	3	4

An employee	1	2	3	4

3. Do any of your immediate family members (parents, siblings) hold an engineering degree?

a. No

b. Yes

c. I prefer not to answer

4. How many classes related to Engineering did you take before college?

a. 0

b. 1

c. 2

d. 3

e. 4 or more

Appendix C

Online Activity Log Release Form

PURPOSE

You are invited to participate in a research study at the University of Tennessee. The objective of the research is to develop an engineering student success model to identify factors associated with engineering students' success and using the model to develop possible intervention to help students who are at risk of failing the course. In order to understand students' learning style and study habits, the researchers are interested in observing students' activities in class and homework studies. The online homework system keeps track of every attempt students have on homework questions. The clicker system keeps track of how students respond to in class questions. The researchers are interested in analyzing students' behaviors on those activities and associating different learning styles with students' learning outcomes.

PARTICIPANT INVOLVEMENT

The researchers request you to release your activity information on Clicker system and Course Management homework system, and your first module test score, to them for analysis. Your name and other identifiable information will not appear on the released data.

RISKS

The risks are minimal to the participants. Participants may take the risk of breach of confidentiality, but the researcher will protect against the risk. The procedures to protect participants' confidentiality are described in the Confidentiality section.

BENEFITS

There is no direct benefit for the participants. The result of the study may help the engineering educator to develop targeted interventions to help first year engineering students to improve their academic success. Further more, the result of the study will provide implications to general engineering curriculum development and modification to better cultivate future engineers.

CONFIDENTIALITY

All of the information will be kept confidential. There will be no names and other identifiable information in the oral or written reports. The data will be saved in a password protected computer and only the researcher can get access to it.

PARTICIPATION

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at anytime without penalty and without loss of benefits to which you are otherwise entitled.

CONTACT INFORMATION

If you have questions at any time about the study or the procedures you may contact the researchers:

Wenshu Li	Dr. Richard Bennett	Dr. Gary Skolits
wli23@utk.edu	rmbennett@utk.edu	gskolits@utk.edu
(951)801-8437	(865)974-7540	(865) 974-6117

If you have questions about your rights as a participant, contact the Office of Research IRB Compliance Officer at (865) 974-7697. Please save a copy of this information for your records.

CONSENT

I have read and understand the above information. I give consent for the researchers to access my homework and clicker data, and my first module test score, in EF 151.

____

Name (please print)

____

Signature Date

Appendix D

Online Survey Consent Form

INTRODUCTION

You are invited to participate in a research study at the University of Tennessee. The objective of the research is to develop an engineering student success model to identify factors associated with engineering students' success and use the model to develop possible interventions to help students who are at risk of failing the course.

It will take your 15-20 minutes to complete the survey and you **will get extra credit for completing the survey.** You will be requested to enter your name in the survey as you consent to participate. The principal investigator (Wenshu Li) will provide your names to your instructor to process the extra credits. If you decide not to participate in the survey, you won't have any grades deduction and you can still complete your homework earlier to earn extra credits.

INFORMATION ABOUT PARTICIPANTS INVOLVEMENT IN THE STUDY

The survey will take you 15-20 minutes to complete, and will request information about your high school and college experience, in particular studying in the College of Engineering at the University of Tennessee, and your demographic information.

RISKS

The risks are minimal to the participants. Participants may take the risk of breach of confidentiality, but the researcher will protect against the risk. The procedures to protect participants' confidentiality are described in the Confidentiality section.

BENEFITS

There is no direct benefit for the participants. The result of the study will help the engineering educator to develop targeted interventions to help first year engineering students to improve their academic success. Furthermore, the result of the study will provide implications to general engineering curriculum development and modification to better cultivate future engineers.

CONFIDENTIALITY

All of the information will be kept confidential. Only the researchers have access to the information you enter. The data will be stored securely in a password required website. There will be no names and other identifiable information in the oral or written reports.

CONTACT INFORMATION

If you have questions at any time about the study or the procedures you may contact the researchers:

Wenshu Li

wli23@utk.edu

(951)801-8437

Dr. Richard Bennett

rmbennett@utk.edu

(865)974-7540

If you have questions about your rights as a participant, contact the Office of Research Compliance Officer at (865) 974-3466.

PARTICIPATION

Your participation in this study is voluntary. You may decline or withdraw to participate in the study at any time without penalty.

CONSENT

I have read and understand the above information. Please save a copy of this information for your records.

Please select your choice below.

Clicking on the "agree" button below indicates that:

• You have read the above information

• You voluntarily agree to participate

• You are at least 18 years of age

If you do not wish to participate in the research study, please decline participation by clicking on the "disagree" button.

- Agree

- Disagree

Your name will be used as a signature of your agreement to participate in this study. The research will use your name and your UTK email address to send you the instruction to earn extra credit and preserve your anonymity.

Your first name______________________

Your Last Name_____________________

Appendix E

Online Survey Recruitment Email

Dear EF151 students,

The Engineering Fundamentals professors are working together with a doctoral student, Wenshu Li, in Educational Psychology department to look for ways to provide additional help and support for students in EF 151. Wenshu has conducted interviews and surveys with EF151 students since 2012 to help us better understand EF151 students' learning and provide recommendations to better engage students' learning in EF151. She would like to invite Fall 2015 EF151 students in her dissertation study, "Factors That Contribute to First Year Engineering students' Academic Success: A structural Equation model". The objective of the research is to develop an engineering student success model to identify factors associated with engineering students' success.

It will take you 15-20 minutes to complete the survey and you **will get extra credit for completing the survey.** You will be requested to enter your name in the survey as you consent to participate. The principal investigator (Wenshu Li) will provide your names to your instructor to give you the extra credits. If you decide not to participate in the survey, you won't have any grades deduction and you can still complete your homework earlier to earn extra credit.

The survey responses will be kept confidential, only the researcher has the access to the survey responses and she will be responsible to keep the survey responses on password-protected computers. We hope that you will agree to participate in the study. We believe that the process will not only benefit you but will also benefit other students in the class by identifying factors associated with your academic success.
Remember, this is completely voluntary. You can choose to be in the study or not.

If you'd like to participate in the study, you can take the survey at the link below:
http://qeasttrial.co1.qualtrics.com/SE/?SID=SV_6MqSEWd8w4PvTYV

If you have any questions about the study, please email or contact Wenshu Li at wli23@vols.utk.edu.

Thank you very much.
Sincerely,
Wenshu Li

Vita

Wenshu Li earned her Bachelor of Science degree in Psychology from Peking University, Beijing, China, where she has been trained in basic statistics and measurement courses. She moved to the United States in 2010 to pursue a master's degree in Educational Psychology at Miami University, Oxford Ohio. Her master's degree at Miami University has a focus on evaluation, statistics, and measurement. She obtained certificates in Statistics and measurement, and program evaluation in the program as well as gained experiences in teaching statistics, field practice in education and public health evaluation, educational research, and survey development and validation.

In 2012 Wenshu Li joined the Evaluation, Statistics, and Measurement doctoral program at the University of Tennessee. During her four years at the doctoral program, she completed a number of evaluation research projects as either principal or co-principal investigator. She has been the graduate research assistant in the College of Engineering to conduct evaluation on k-12 and college engineering education program applying her expertise in statistics, measurement, and evaluation. Moreover, she also has been disseminating knowledge through various research, evaluation, and grant reports, and presentations at professional conferences. While Wenshu pursing her PhD degree in the ESM program, she also completed a master's degree in Statistics at the University of Tennessee. She worked with Dr. Wenjun Zhou, who is her advisor in the program on several educational data mining projects. These experience as well as her graduate assistant experience in the college of engineering has led her to conduct her dissertation study. Wenshu Li graduated from the University of Tennessee, Knoxville in May of 2015 with a PhD in Educational Psychology and Research with a concentration in Evaluation, Statistics, and Measurement. As of Auguest 2015, Wenshu has taken a post-doctoral evaluation fellowship position at the Center of Disease Control and Prevention.