Author: **Johnson, Michael A.**

Title: ***An Analysis of Student Persistence at Mid-State Technical College***

The accompanying research report is submitted to the **University of Wisconsin-Stout**, **Graduate School** in partial completion of the requirements for the

Graduate Degree/ Major: MS in Career and Technical Education

Research Advisor: Dr. Brian McAlister, Teaching, Learning and Leadership Dept.

Submission Term/Year: WinTerm 2019

Number of Pages: 47

Style Manual Used: American Psychological Association, 6th edition

☒ **I have adhered to the Graduate School Research Guide and have proofread my work.**

☒ **I understand that this research report must be officially approved by the Graduate School.**

☒ **I attest that the research report is my original work (that any copyrightable materials have been used with the permission of the original authors), and as such, it is automatically protected by the laws, rules, and regulations of the U.S. Copyright Office.**

☒ **My research advisor has approved the content and quality of this paper.**

STUDENT:

NAME: Michael A. Johnson **DATE: 1/14/2019**

ADVISOR: (Committee Chair if MS Plan A or EdS Thesis or Field Project/Problem):

NAME: Dr. Brian McAlister **DATE: 1/14/2019**

This section for MS Plan A Thesis or EdS Thesis/Field Project papers only

Committee members (other than your advisor who is listed in the section above)

1. **CMTE MEMBER'S NAME:** **DATE:**
2. **CMTE MEMBER'S NAME:** **DATE:**
3. **CMTE MEMBER'S NAME:** **DATE:**

This section to be completed by the Graduate School

This final research report has been approved by the Graduate School.

Director, Office of Graduate Studies: **DATE:**

Johnson, Michael A. ***An Analysis of Student Persistence at Mid-State Technical College***

Abstract

This study analyzed student persistence at Mid-State Technical College by first measuring the accuracy of the predictions made by a newly acquired software that uses available student data to predict how likely they are to persist. To do this, the software's prediction scores for degree seeking students at the college were captured every two weeks throughout the Spring 2018 semester. These prediction scores were then compared to actual student persistence rates (those that either enrolled in the Fall 2018 semester or graduated). The accuracy of the software's predictions were measured by calculating the correlation coefficient (R^2) between the software's prediction scores and actual student persistence at the college. R^2 values above 0.95 were considered very strong, with values above 0.90 still considered strong. An R^2 value below 0.90 was considered to be a weak correlation.

Once the software's accuracy was determined, further analyses compared persistence among different demographic groups and programs at the college to identify areas of opportunity for improved student success.

Acknowledgments

First, I must thank my wife Kim for supporting me through the pursuit of this degree and the culminating research contained in this paper. Without your love and support, none of this would have been possible. Managing our chaotic house of four young children (two of which were born during my pursuit of this degree) and continuing to work and grow in your own career while I spent many long nights and weekends reading, studying, and researching in an effort to earn a second master's degree was no small task. I could not have done it without you.

Second, I must thank my employer, Mid-State Technical College, for giving me the opportunity to pursue this degree, and in so doing, become a better instructor. "Mid-State Technical College transforms lives through the power of teaching and learning." This mission statement is at the heart of what we do, and your investment in me as a faculty member will help me better serve our students and promote their success.

Lastly, I must thank the many teachers I have had the pleasure of learning from. Teaching can sometimes be a thankless, daunting, and exhausting profession. Inspiring students to learn something entirely new requires an incredible amount of patience. My formal education began in kindergarten more than thirty years ago, and every teacher that has worked with me during this time has contributed in some way to the success I enjoy as a college instructor today. While I may not have thanked each of you as our paths diverged, I now realize how empowering a simple thank you from a graduating student can be.

Table of Contents

List of Figures

Chapter I: Introduction

Of the numerous challenges facing higher education in America, student persistence is among the most talked about, analyzed, scrutinized, and important. The national Career and Technical Education research agenda guides research efforts toward five problem areas: Accountability, Program Relevance and Effectiveness, Knowledge Base for Teaching and Learning, Curricula and Program Planning, and Delivery Methods (Lambeth, Joerger, & Elliot, 2009). Within this framework, student persistence best aligns with the area of *Accountability*, and as such, is a valuable topic to consider.

Countless studies have explored the issue, considering both why students leave college and how to keep them from doing so (Peltier, Landen, & Matranga, 2000). Common variables that can help predict student persistence include gender, age, year in college, socioeconomic status, first generation college student status, ACT score, and high school rank (Maas, 2008), of which educational and financial variables are often considered most predictive (Delen, 2011).

Promoting student success is at the core of what any academic institution aspires to accomplish. Given the level of attention devoted to this topic, one would expect related outcomes to improve over time. Unfortunately, they have not. Since 2009, persistence among first-year college students attending two-year public institutions full-time has remained at or near 70% (National Student Clearinghouse Research Center [NSCRC], 2017). Those attending part-time have persisted at a rate of approximately 55% during this period. When considering similar students at a four-year public institution, full-time students persisted at or near a rate of 88%, while part-time students persisted at or near a rate of 62% over this period (NSCRC, 2017). Despite efforts to improve outcomes, persistence rates within each group remained rather consistent.

While this apparent lack of improvement may be discouraging, the reasons behind it are both numerous and complex. One important variable to consider is the college's selectivity. Colleges that accept less than 25% of applicants retain 96% of their first year students, whereas colleges with open admission models retain approximately 60% of first year students (National Center for Educational Statistics, 2017). This disparity is apparent when comparing graduation rates as well (Sullivan, 2008). These data show how implementing open access enrollment policies may put downward pressure on student persistence rates by allowing students to begin programs for which they are not adequately prepared.

To the contrary, a recent study on this topic found that the average student, scoring 1000 on the SAT exam, would be more likely to graduate if they chose to attend a selective university than if the same student chose to attend an open access college. Despite greater rigor, the average student persists at a higher rate (Strahota, 2016). This seems counterintuitive, but hints at the complexity of the student persistence issue.

Improving student persistence is an ongoing effort at many colleges, including Mid-State Technical College in central Wisconsin. While 70% of students who began attending full-time at Mid-State in the fall of 2014 returned the following year, only 31% of those beginning their program in the fall of 2013 graduated after three years, or 150% of the normal time-to-completion (Mid-State Technical College, 2018).

In an effort to improve these metrics, Mid-State is implementing the use of *Illume Students*, a data analysis software offered through Civitas Learning that predicts the likelihood of student persistence using available student data (Civitas Learning, 2018). It is anticipated that these predictions will help the college prioritize student service efforts toward those predicted to be at risk.

Statement of the Problem

Mid-State Technical College aspires to promote student persistence to better serve its students. The problem is that student persistence is a complex and elusive target, and despite recent efforts, persistence rates at the college remain below desired levels. Absent a thorough understanding of the students in question, improving their persistence continues to be a challenge. In order to gain this understanding, and ultimately use it to improve student persistence, the college has implemented the use of *Illume Students* software.

How useful the software is to the college relies on the accuracy of the predictions that the software makes. As valuable as accurate predictions of student persistence would be to the college, inaccurate predictions could be misleading and detrimental.

Purpose of the Study

The purpose of this study was to determine the usefulness of the recently implemented *Illume Students* software at Mid-State Technical College. More specifically, the study analyzed the accuracy of its predictions and explored what patterns exist at the college regarding student persistence. With this information, resources can be better allocated toward improving student persistence at Mid-State Technical College.

This research helped the college determine the software's usefulness as a tool to improve student persistence at Mid-State Technical College. Once the accuracy of the predictions was determined, further analysis compared persistence among different demographic groups and programs over the course of the Spring 2018 semester. Identifying student groups that persist at lower rates, as well as critical times within the semester when students tend to falter will give the college a better understanding of its student body. This understanding will help guide concentrated efforts toward those at risk to improve student persistence at Mid-State Technical College.

This research addressed the following questions:

1. How accurately does the Illume Students software predict student persistence at Mid-State Technical College?
2. How does student persistence vary over time at the college?
3. At what rate do different demographic groups at the college persist?
4. At what rate do students from different schools within the college persist?

Assumptions of the Study

The underlying assumption for this study was that students pursuing a degree at the college intend to persist and ultimately earn their chosen degree. While this is generally true, it may not always be so. Identifying and excluding those that do not intend to persist is very difficult, if not impossible. For that reason, it was assumed that all degree-seeking students at the college aspire to persist and complete their degree, and efforts by the college to promote their success are warranted.

Definition of Terms

The terms below are defined as they relate to this paper and the research performed herein.

Accuracy. For the purposes of this study, accuracy was measured by calculating the correlation coefficient (R^2) between the software's persistence prediction score and the actual persistence rates of students.

At risk. For the purposes of this study, those students predicted to have a "Very Low" or "Low" likelihood to persist (persistence score below 50%) were considered to be at risk.

Full-time students. For the purposes of this study, a student was considered full-time if they were enrolled in twelve or more credits in a given semester.

Part-time students. For the purposes of this study, a student was considered part-time if they were enrolled in fewer than twelve credits in a given semester.

Persistence. For the purposes of this study, a student demonstrated persistence by being a current student at Mid-State Technical College, and enrolling in the subsequent semester at Mid-State Technical College, or by graduating.

Persistence prediction. The output of the Illume Students software, which rates a student's likelihood to persist as "Very Low" (prediction score of 0%-20%), "Low" (prediction score of 20%-50%), "Moderate" (prediction score of 50%-70%), "High" (prediction score of 70%-90%), or "Very High" (prediction score of 90%-100%).

Persistence prediction score. Percent score produced by the software that estimates how likely a student is to persist. Prediction scores range from 0% to 100%.

School. The programs offered at Mid-State Technical College are organized into six "schools" which served as secondary independent variables in this study.

Limitations of the Study

Limitations of the study include the following:

1. This study considered students enrolled at Mid-State Technical College only, and as such, its conclusions are not necessarily applicable to other colleges using the software.
2. The data in the study considered persistence predictions/results from the Spring 2018 to Fall 2018 semesters. This abbreviated timeline allowed timely feedback to the college, but limited the depth of the analysis.

Methodology

This research used an ex post facto methodology to gather data on a variety of variables. The independent variables in this study were the predictions made by the *Illume Students* software, as well as the student's program/school and demographic category. The dependent variables included the accuracy of the predictions made by the software and the persistence rates of students in each group. The control variable in this study was time, as all data was gathered between the Spring 2018 and Fall 2018 semesters.

Chapter II: Literature Review

The following review of literature synthesized what is currently known regarding student persistence in higher education, and gives context to the study at hand.

Theory Base

Student persistence has been a topic of research for decades. Some studies seek to identify why students leave college, others investigate strategies to improve student persistence, and still others explore what characteristics are common in students who depart vs. persist, including both demographic and behavioral attributes. One such attribute is self-efficacy, or a student's confidence regarding given tasks. Those that are confident in their abilities are more likely to persist than those with considerable doubt (Wright, Jenkins-Guarnieri, & Murdock, 2012). While this may be true, self-efficacy is a trait of the incoming student, is difficult to impart, and is therefore of limited use to an institution hoping to improve persistence of an existing student body.

While many label students who do not persist as "dropouts" and assume that the rigor of the college classroom caused their departure, the reality is that students leave college for a variety of reasons, many of which relate to the college itself as opposed to the student. Understanding the reasons why students leave college is fundamental to improving their persistence.

The considerable volume of research on this topic has led to numerous theories that attempt to explain it. One of the most widely accepted is Vincent Tinto's Theory of Student Departure, which argues that colleges are like other human communities: the health of which is a dynamic byproduct of the daily interactions of its members (Tinto, 1999). Tinto's theory details three stages a student must navigate to fully integrate into a college community: separation, transition, and incorporation. In essence, the student must first separate from prior communities (high school, friends, and family), transition toward the new college community with different norms and behaviors, and become incorporated into the new community by adopting these norms to ultimately become integrated into the new community (Milem & Berger, 1997). Given this, it follows that "Decisions to withdraw are more a function of what occurs after entry than of what precedes it" (Tinto, 2012, p. 5).

While this seems to ignore correlations between existing student characteristics and persistence (Maas, 2008), it shifts attention away from the shortcomings of students and toward what the institution can do to foster a healthy "community" that promotes persistence of all student members. It is because of this theory's proactive approach that it serves as the base for the following research.

Illume Students Software

The primary independent variable in question was the predictions made by the *Illume Students* software, which uses available data to predict a student's likelihood to persist. While the software's prediction algorithm is proprietary, and therefore unavailable for analysis, a unique list of "powerful predictors" was given within the software for the group of students (i.e. male, female, full-time, part-time, etc.) being considered under the active filter. These predictors "…use historical data to show what variables are important to persistence for this group of students." (Civitas Learning, 2018). These lists included over one hundred items and varied widely depending on the filter being applied, but common predictors among filters included degree being pursued, high school and college GPA, credits earned at the institution, change in GPA, marital status, course load, financial status, age, gender, race, number of days enrolled before start of term, etc. Many of the student groups analyzed in this paper appeared on the powerful predictor lists, implying that they are important factors to consider regarding persistence.

Tinto reasons that "…while it is important that universities challenge existing labels as marking some entering students as less likely to succeed than others, it is equally as important that students are able to obtain the timely support they need when they encounter early difficulties…" (Tinto, 2017, p. 3).

The college intends to use the *Illume Students* software to help focus retention efforts toward those predicted to be at risk. This software was used to gather all data analyzed in the study.

Student Demographic

One secondary independent variable was the demographic of the students in question. Demographic groups analyzed for persistence rates included race, gender, full-time vs. part-time status, and financial aid vs. self-paid students.

Student Program/School

Another secondary independent variable was the different schools within the college. The programs offered at Mid-State are organized into six "schools." These include the School of Health, School of Advanced Manufacturing and Engineering, School of Protective and Human Services, School of General Education and Learning Resources, School of Business and Technology, and School of Transportation, Agriculture, Natural Resources, and Construction.

Prediction Accuracy

The primary dependent variable in this study was the accuracy of the predictions made by the software. Civitas Learning, the manufacturer of *Illume Students* software, details numerous success stories of colleges using their products. They also stress that *Illume Students* software is not a “one size fits all” product, but rather, employs a tailored algorithm using the college’s own data to predict how likely each student is to persist. If necessary, the client can work directly with Civitas to adjust this algorithm in an effort to improve its accuracy (Civitas Learning, 2008).

Verifying the accuracy of these predictions was a critical first step, and was done by calculating the correlation coefficient between the software’s persistence prediction score and the actual persistence rates of students.

Student Persistence

The secondary dependent variable in this study was the rate of persistence of students within each school and demographic group. Numerous studies have demonstrated measurable differences in persistence rates among these demographic groups. For instance, one such study analyzing data from over five hundred universities found that Asian students persisted at a rate of 61.1%, compared to white students persisting at 56.9%, while Black (41.7%), Hispanic (41.7%), and Native American (35.8%) students persisted at significantly lower rates (Seidman, 2005).

Another study found that one of the most powerful predictors of student persistence is the socioeconomic status of the student’s parents (Delen, 2011). Comparing these persistence rates across different student groups uncovered areas of opportunity for improving persistence at the college.

Summary

Given the importance of student persistence in higher education, it should be no surprise how thoroughly the topic has been researched. Among those researching student persistence, Vincent Tinto is considered an expert, and his Theory of Student Departure continues to be widely accepted as an explanation of why students persist in college. Grounding this research in

Tinto’s theory provides a framework for exploring a critical and complicated part of higher education.

Chapter III: Methodology

The purpose of this study was to verify the accuracy of the predictions made by the newly acquired *Illume Students* software at Mid-State Technical College. The software was then used to gather data in order to analyze student persistence at the college. More specifically, persistence was explored to determine if it varied among different schools and demographics within the college.

Subject Selection and Description

The subjects for this study were all of the degree-seeking students attending Mid-State Technical College between the Spring 2018 and Fall 2018 semesters. Data for these subjects was readily available through the software, and was therefore a convenient sample. While this sample group was quite large, it did not include all students attending Mid-State Technical College. Some students at the college were not seeking a degree, but rather pursuing continuing education. As such, these non-degree seeking students were not expected to persist to the ensuing semester, and therefore were not included in this study. Given the size of the degree-seeking sample, the results of the study are generalizable for future degree-seeking students attending Mid-State Technical College. Conversely, results are not generalizable to students attending other post-secondary educational institutions.

Instrumentation

The instruments used for this study included screen shots from the overview page of the *Illume Students* software (Figure 1 below), as well as active student lists (Figure 2 below).

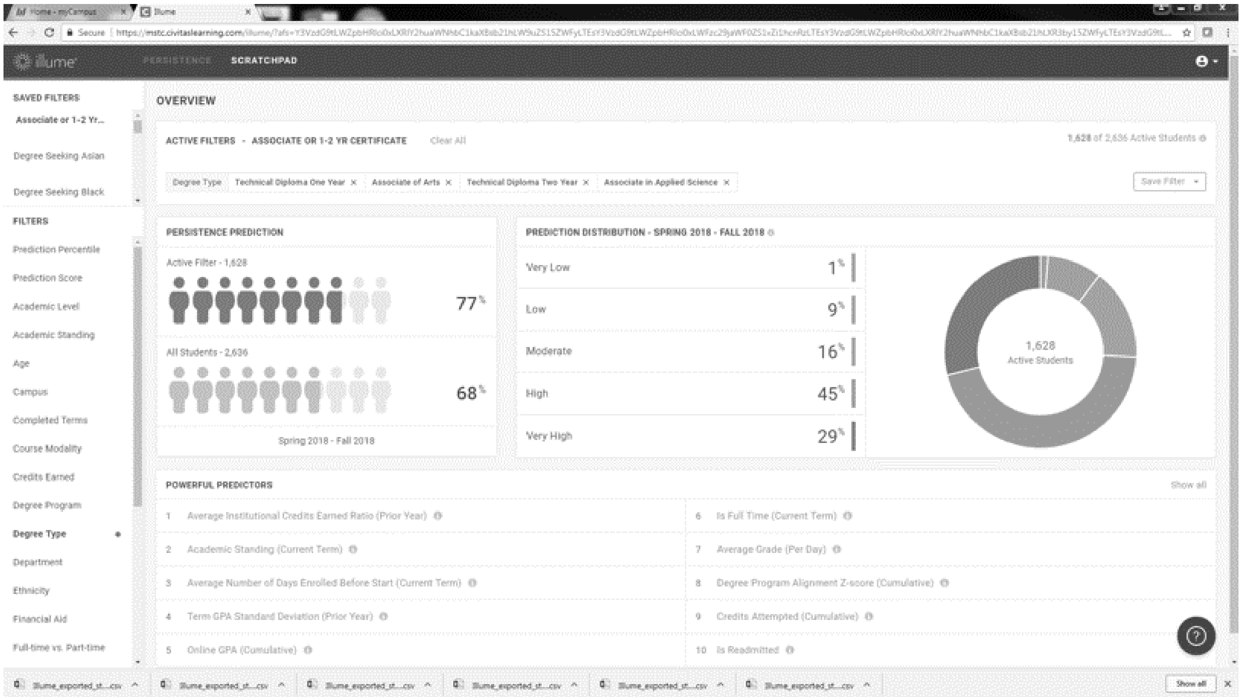


Figure 1. Screenshot of software prediction summary for degree seeking students.

Student ID	First Name	Last Name	Email	Enrolled Current Term	Persistence Prediction	Persistence Prediction Score	Last Enrolled Term	Next Enrolled Term
xxx	xxx	xxx	xxx	Yes	Low	43.07%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	Moderate	69.58%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	High	88.71%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	High	84.75%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	Moderate	64.96%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	Very High	93.72%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	High	84.16%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	Moderate	64.67%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	High	83.64%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	Moderate	66.45%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	Moderate	61.88%	Spring 2018	-
xxx	xxx	xxx	xxx	Yes	High	87.33%	Spring 2018	Summer 2018

Figure 2. Active student list data for degree seeking students.

Data Collection Procedures

The software's predictions were captured every two weeks throughout the study's duration by taking screen shots from the software interface, as well as exporting active student list spreadsheets. This periodic documentation was necessary because past predictions are not recoverable at a later date in the software. Filters within the software categorized the students (and their corresponding persistence predictions) into comparable groups, including race, sex, financial aid status, full-time vs. part-time status, and the school at the college to which their chosen program belongs. The steps taken to gather data were as follows:

1. At the beginning of the study, filters were established within the software that included race, sex, financial aid status, full-time vs. part-time status, and program school.
2. Data collection began during the third week of the Spring semester by applying these filters and taking screen shots of the software's interface (it was not possible to generate a report for the overall predictions within the software), as well as exporting active students lists for each filter. Data collection began in the third week because the college's "census" date occurred in week two, after which students could no longer enroll in a class that had already begun. Additionally, students could drop a class and receive a full refund of their tuition if done so before 10% of the class meetings had occurred. For full-semester courses, this threshold occurred during the second week of the semester. For these reasons, collecting data prior to week three likely would have generated erroneous results, as students schedules were still in flux.
3. Data collection continued every two weeks until the start of the Fall 2018 semester to document how the data changed over time.
4. When the Fall 2018 semester began, active student lists and screen shots were captured within the software to summarize actual enrolled students. Additionally, a Spring 2018 graduation list determined which of those no longer enrolled persisted

via graduating. These two data sets were then combined and compared to the persistence predictions collected throughout the Spring semester.

Data Analysis

Simple descriptive statistics were employed to analyze the data in this study. The accuracy of the predictions made by the software was measured by comparing the prediction scores produced by the software to reality, and calculating the correlation coefficient (R^2) between these scores and actual persistence. This correlation was measured using a linear regression analysis to determine an R^2 value between the mean prediction scores within each category (Very Low, Low, Moderate, High, Very High) and actual persistence rates.

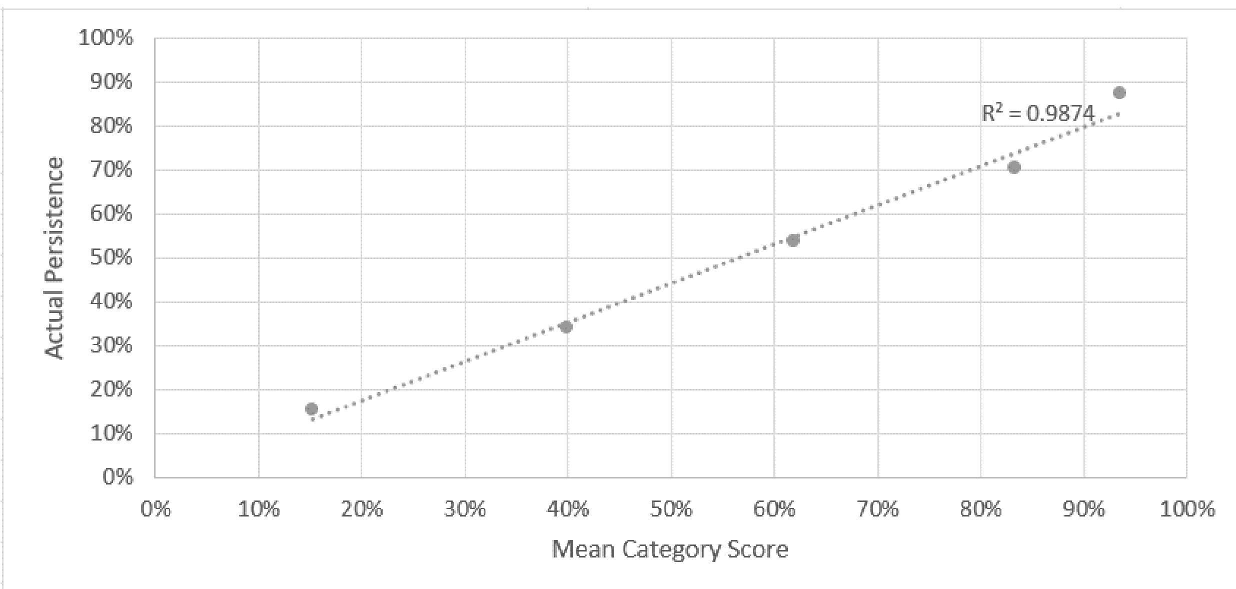


Figure 3. Example of graph used to produce R^2 values.

These prediction categories only provided five data points from which to calculate R^2 values, whereas creating finite ranges of predictions scores (0%-5%, 5%-10%, etc.) would have generated more data points for this calculation. The study used the five prediction categories rather than smaller finite ranges because most end users of the software view these prediction categories from a summary page (Figure 1) rather than generate an active student list that includes actual persistence scores (Figure 2). Choosing to use the five categories produced a better reflection of the software's accuracy as it is used by faculty and administration at the college.

R^2 values were calculated for each collected set of predictions (every two weeks throughout the Spring semester) to determine times that prediction accuracy fluctuated. R^2 values above 0.95 were considered very strong, with values above 0.90 still considered strong. An R^2 value below 0.90 was considered to be a weak correlation. If the correlation was strong or very strong, then the predictive power of the software was considered reasonably accurate. If this correlation was weak, then the predictive power of the software was questionable, and the algorithm that the software used to make its predictions was likely not considering important factors.

Summary

The methodology described above captured predictions that change over time and are not recoverable at a later date within the software being analyzed. Determining to what extent these predictions were realized was the impetus of this study. Once the accuracy of the software was determined, further analysis of how these predictions changed over time among the independent variables of the study was performed and presented in graphical format to identify opportunities to improve persistence at the college.

Chapter IV: Results

With the data collected and compiled into spreadsheets and graphs, the research questions posed in this study were addressed as follows:

Research Question 1: How Accurately Does the Illume Students Software Predict Student Persistence at Mid-State Technical College

The answer to this question was sought by measuring how closely the mean prediction score within each category (Very Low, Low, Moderate, High, Very High) correlated with actual student persistence, and how this correlation varied over the course of the semester. When considering all degree seeking students at the college (over 1,700 students), the R^2 value was above 0.95 for all but two weeks of the semester. The only deviation from this trend was a slight drop as the end of the semester approached; weeks fifteen and seventeen.

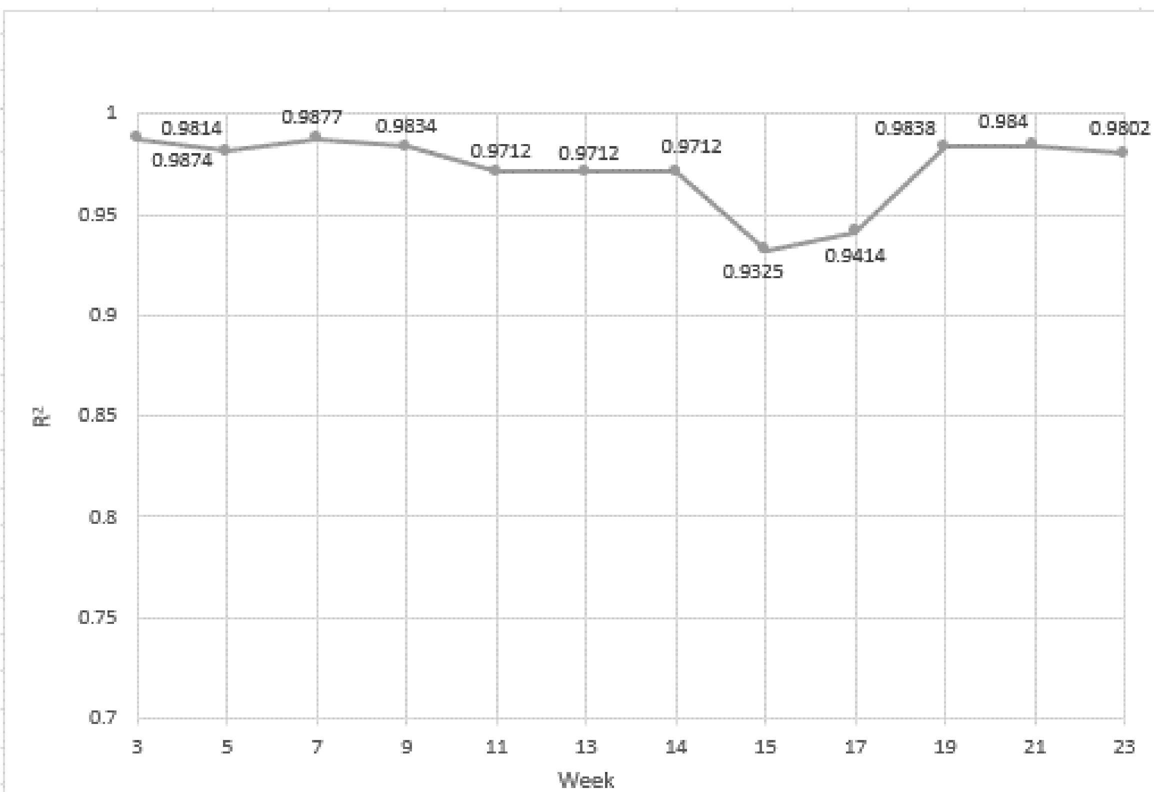


Figure 4. R^2 vs. time (degree seeking students).

This drop was only slight, with R^2 values remaining above 0.90, the threshold for what the study considered to be a strong correlation.

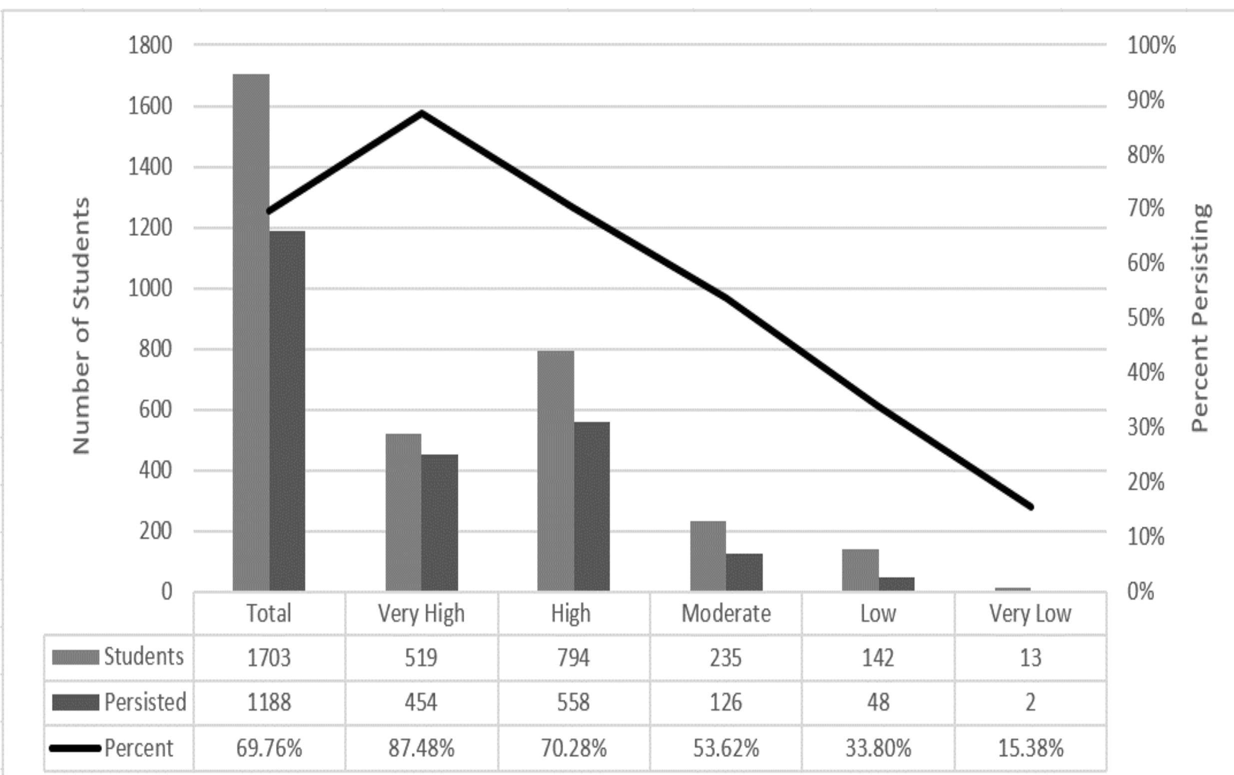
The timing of this decline seems to agree with the deadline to drop a course and receive a W grade rather than and "F". This deadline is before 90% of the course is complete, which occurred during week sixteen for a class that ran the entire semester.

Another explanation of this decline might be that a large portion of a student's grade is often determined at or near the end of the semester, with final projects and exams coming due. These looming deadlines may cause struggling students to either withdraw from a given class or even the college entirely. The software's persistence prediction algorithm likely saw nothing to indicate this change in status until it actually occurred, creating lower correlation values at the end of the semester. Once these changes occurred and the software's predictions adjusted accordingly, correlation values improved to very strong.

While these results seem to indicate that the software predicts student persistence with reasonable accuracy, similar correlation analyses within each demographic group and school at the college produced somewhat erratic R^2 values, many below the 0.90 threshold the study considered to be a strong correlation. Upon closer inspection, however, it was apparent that as the sample size increased, so did the R^2 value. Groups small enough to only have a handful of students in any given prediction category (Very Low, Low, etc.) typically produced a wider range of R^2 values. This was not surprising, as the persistence outcome of one student in such a category can produce far higher or lower actual persistence within the category than was predicted by the software.

Based on the R^2 values among the different groups/sample sizes, it appeared that strong or very strong correlation was typically found if the group being analyzed had at least one hundred students in it, and each prediction category within the group had a minimum of 15-20 students. If both of these conditions were not met, the correlation became inconsistent, and was sometimes found to be weak (less than 0.90). The groups analyzed in this study that did not meet the minimum sample size of one hundred were the minority races, while several other groups did not meet the minimum number of students per prediction category at some point during the semester. In either case, R^2 values became erratic, and sometimes trended lower in these instances.

That being said, the software's accuracy become apparent when inspecting Figure 5 below, which shows a linear relationship between the percent persisting among the different prediction categories for degree seeking students at the college.



	Total	Very High	High	Moderate	Low	Very Low
Students	1703	519	794	235	142	13
Persisted	1188	454	558	126	48	2
Percent	69.76%	87.48%	70.28%	53.62%	33.80%	15.38%

Figure 5. Persistence per category (degree seeking students).

As the category changes from Very High to High, Moderate, Low, and Very Low, the percent of students persisting within each category decreased in a very linear fashion.

This phenomenon can be seen in a different format when looking at Figure 11 on page 33, which shows that the actual persistence rates of students in each category were within or near the persistence scores that defined said category.

Again, some turbulence was seen near the end of the semester (week fourteen through seventeen), but the lines for each category did not deviate far from their predicted persistence range, and did not intersect each other.

Having explained anomalies in R^2 values within small groups, and demonstrating consistency among large sample sizes, it appears that, depending on the sample size, the software predicted student persistence reasonably accurately. That being said, further analysis of persistence among different demographics and schools within the college was warranted.

Research Question 2: How Does Student Persistence Vary Over Time at the College

The intent of this question was to determine if critical times during the semester exist that the college can target to improve student persistence. To answer this question, data was collected as the semester progressed. This data was then compiled and used to produce Figures 6 through 11, each of which shows how a given data set varies over time.

Figure 6 shows how the number of degree seeking students and their actual persistence rates varied over time.

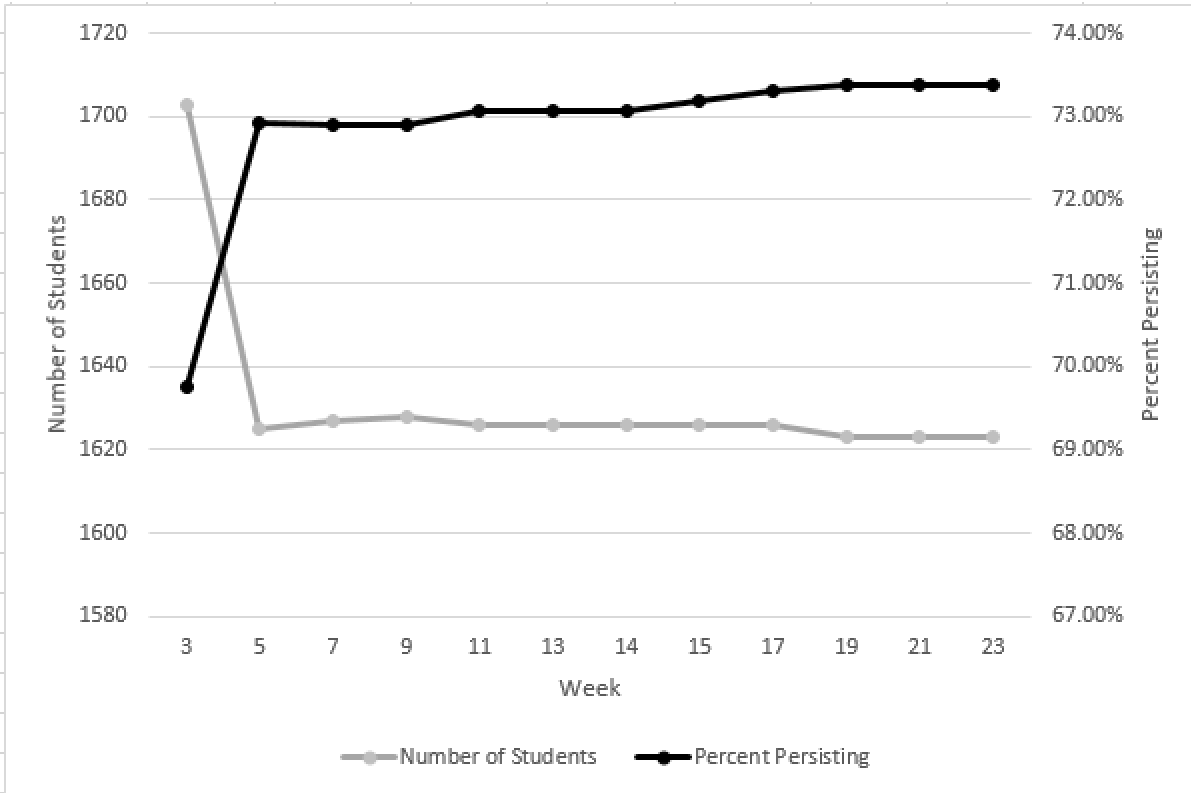


Figure 6. Number of students and persistence vs. time (degree seeking students).

Two things are apparent when looking at this graph. First, the number of degree seeking students dropped drastically before week five, then remained rather consistent thereafter. Second, the percent of students persisting had a measurable increase between week three and five, and then gradually trended upward thereafter. While this is not surprising, it does underscore the importance of the early weeks in the semester. Many of students who did not persist withdrew before week five, and those that remained enrolled after week five persisted at higher rates (72.92% vs. 69.76%).

Figure 7 compared average persistence prediction scores and actual persistence rates vs. time.

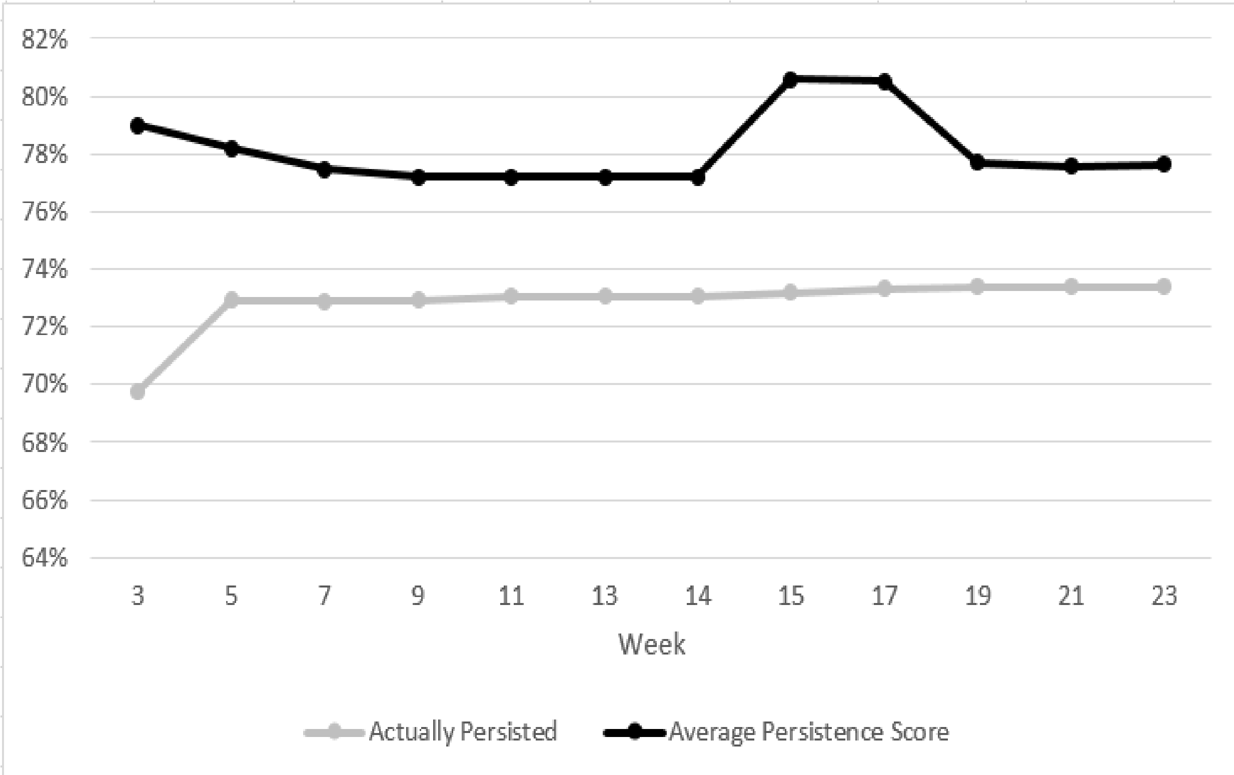


Figure 7. Average persistence score and actual persistence vs. time (degree seeking students).

One thing that was apparent in the above figure was the increase in average persistence score between weeks fourteen and seventeen. A possible explanation for this increase was the fact that enrollment for the following term opened at the end of week thirteen. While many students enroll on this date, some wait several weeks to enroll for the following term, as they may be unsure of their plans moving forward, have holds on their accounts, etc. One would expect an increase in predictions scores when students enroll for their next term, as most doubt regarding their intentions has been removed. One would expect to see the increase happen as soon as enrollment opens at the end of week thirteen, and therefore show up in the week fourteen data point. The fact that the increase in average persistence scores did not happen until week fifteen suggests that either enrollment opening was not the cause of the increase, or there was a lag of approximately two weeks between enrollment opening and scores reflecting this new information.

The decline in average persistence scores after week seventeen (the last week of the semester) was likely due to the end of the semester drop date discussed earlier, and if so, a similar two week lag appears to exist between this time and its reflection in the software's predictions.

Figure 7 also underscores how important the first five weeks of the semester were regarding student persistence, with actual persistence rates increasing until week five and remaining rather constant thereafter. Conversely, average persistence scores trended downward during this time, stabilizing in week seven. This timeline supports the idea that there is a two week lag between reality and the software's predictions, and also indicates some inaccuracy in its early predictions. Had the students withdrawing from classes in the early weeks of the semester been predicted to do so by the software, the average persistence score likely would have increased during this time as students with lower prediction scores withdrew.

While changes in prediction scores relative to identifiable new information indicated a level of accuracy regarding the software's algorithm, the two week lag is concerning. Identifying and helping at risk students is extremely time sensitive; offering them guidance two weeks after their dilemma takes place is often too late to be effective.

A final item to note in Figure 7 is that the average persistence score remained higher than the actual persistence rate across the entire semester. This was the case in every student group analyzed in the study except among Native American students. When considering mean persistence scores, which were higher than average scores in every case, the Native American group had scores higher than actual persistence, similar to every other group analyzed.

Further evidence to that point is apparent in Figures 8, 9, and 10, which show the vast majority (nearly 80%) of degree seeking students at the college were considered Highly or Very Highly likely to persist.

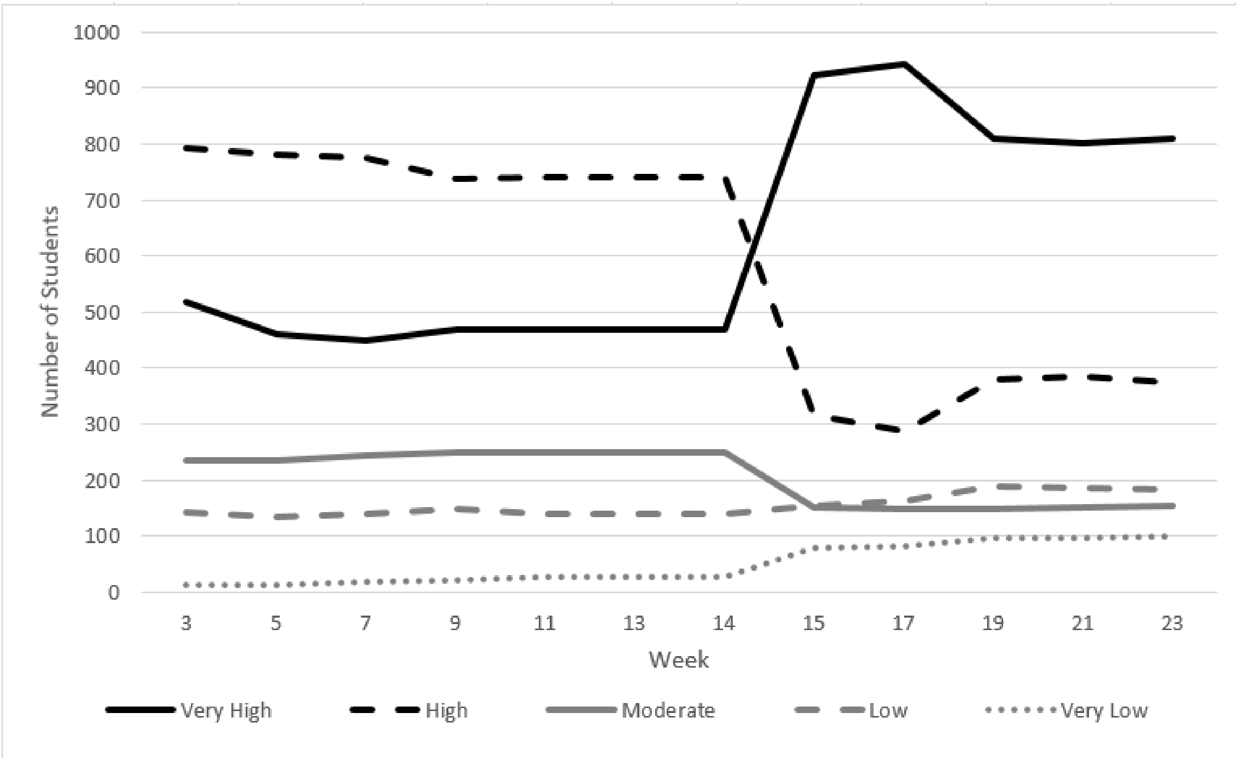


Figure 8. Number of students per category (degree seeking students).

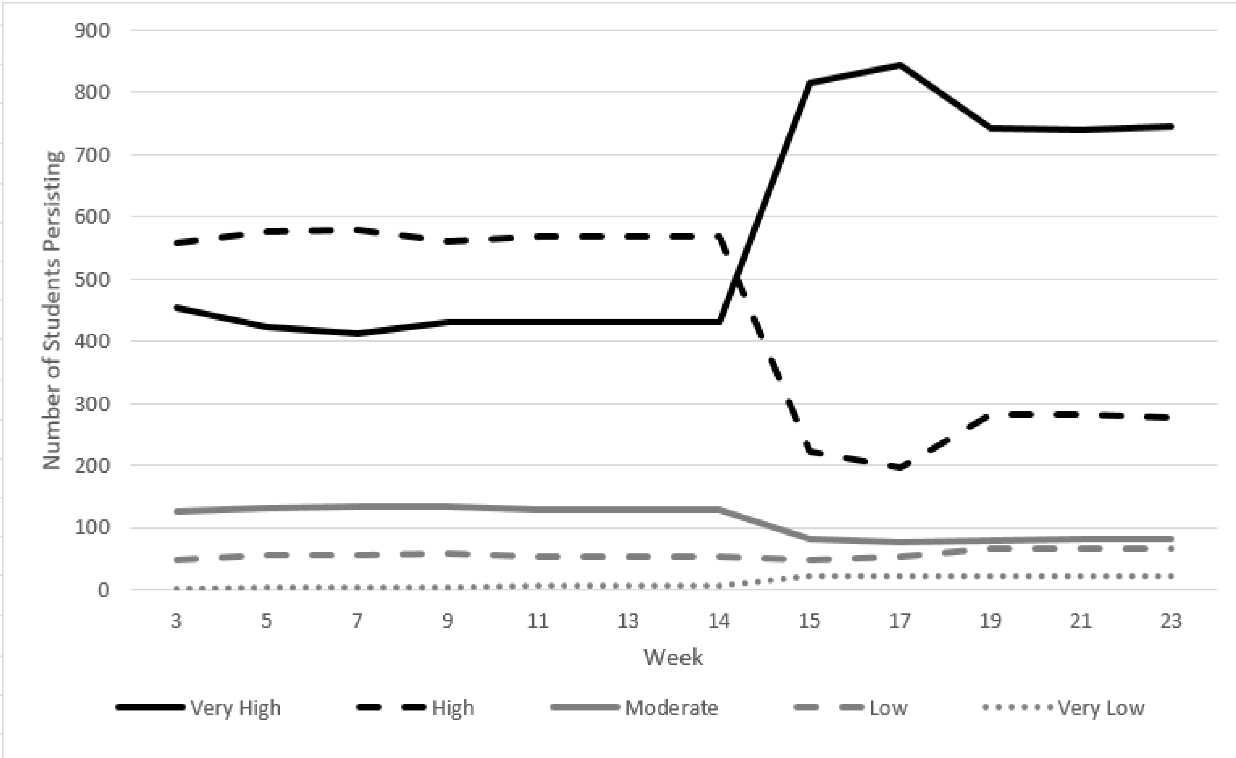


Figure 9. Number of students persisting in category (degree seeking students).

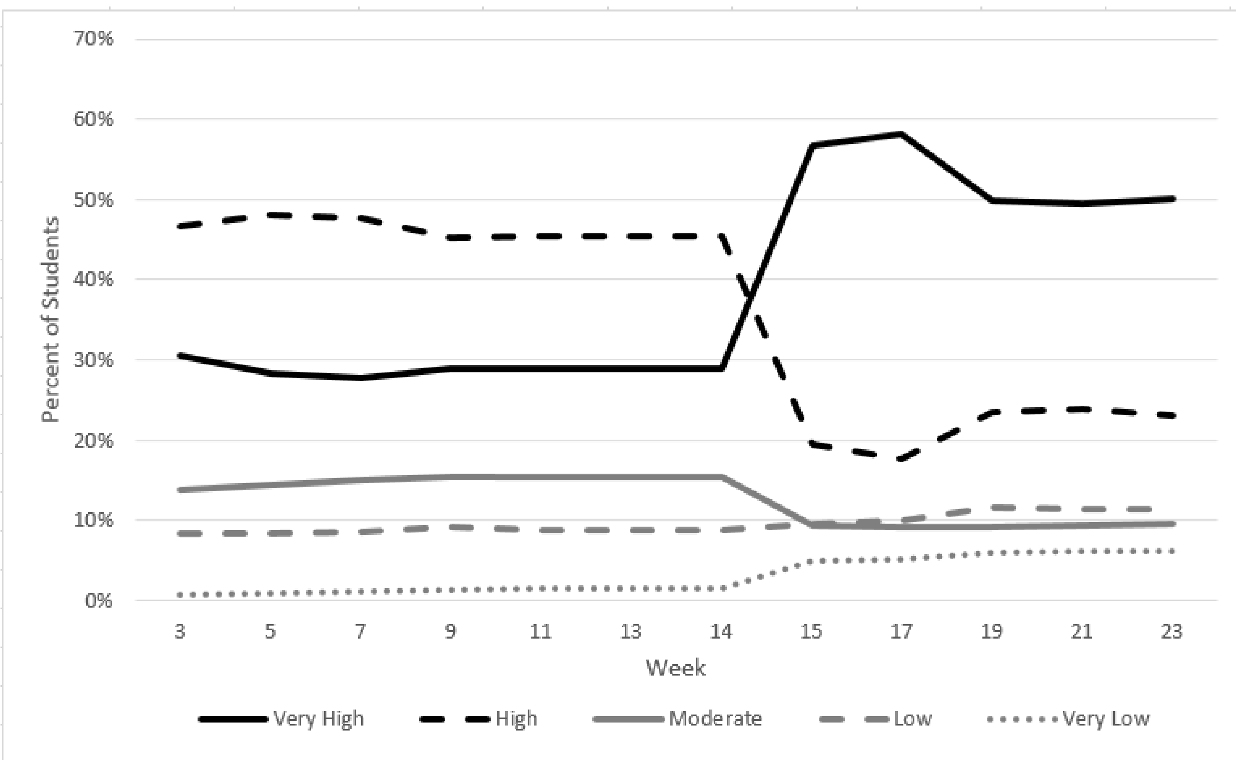


Figure 10. Percent of students in each category (degree seeking students).

Additionally, the mean persistence score for degree seeking students at the college in week five was 84.12%. Considering the fact that the actual persistence rate among degree seeking students at the college was approximately 70%, it appears that these prediction scores were either overly optimistic (not considering important inputs, unable to predict personal circumstances of the student, etc.), or not a direct prediction of persistence rate. This is not to say that the software was inaccurate, but rather, that one should not interpret the prediction scores as expected actual persistence rates.

Another notable trend present in Figures 8, 9, and 10 is turbulence in the graphs between weeks fourteen and nineteen. This was seen regarding R^2 values in Figure 4, and was likely due to the same reasons discussed earlier (enrollment and withdrawal dates, final projects, exams, etc.). This turbulence seems to indicate inaccuracies in the predictions of the software. To the contrary, however, one could argue that most of the reclassification of students during this time period appears to happen between the High and Very High category, indicating that the predictions were relatively accurate, and in light of newly available information, needed only slight modification.

To a lesser extent, it appears that a proportionate number of students were reclassified from the Moderate to the Very Low category during this time. Again, it could be argued that these are minor, not drastic, shifts in predictions based on newly available information as enrollment and drop dates occurred two weeks prior.

That being said, Figure 11 does not seem to support this argument.

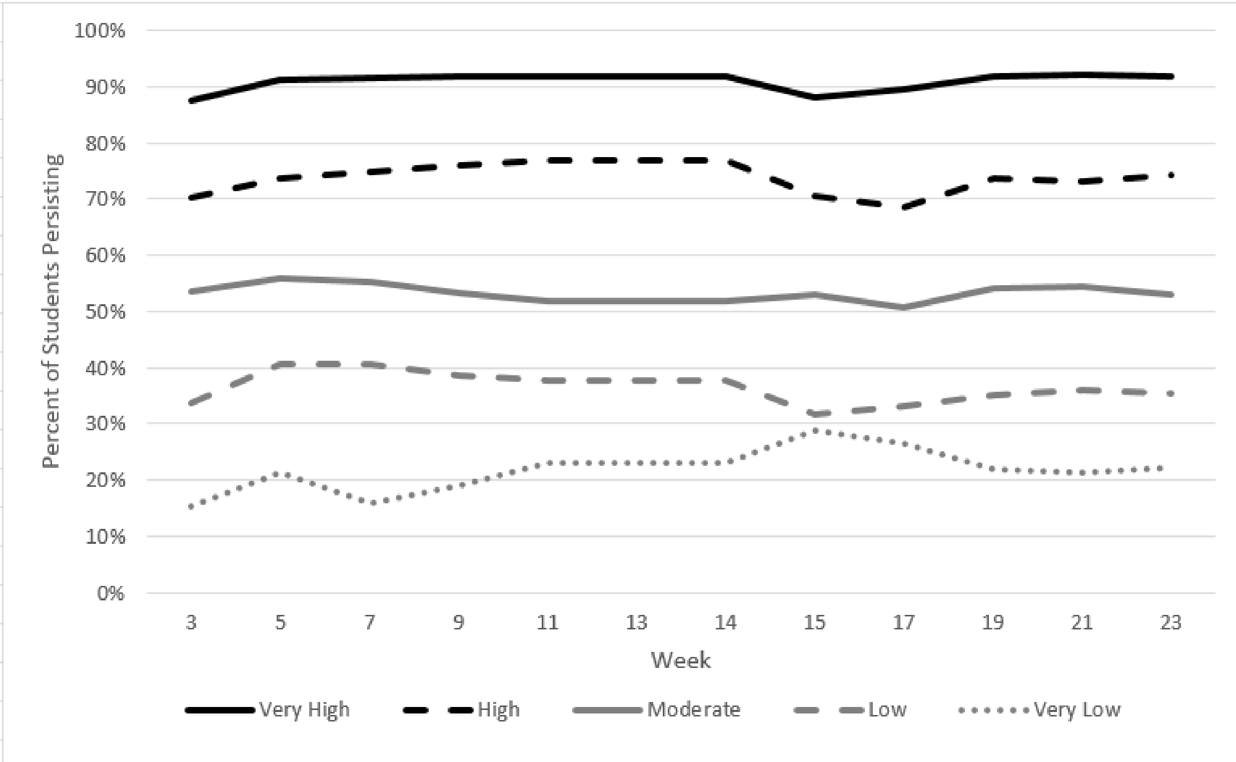


Figure 11. Percent of students actually persisting in each category (degree seeking students).

If students were being downgraded from the Very High to the High category, one would expect the percent of students persisting in both categories to increase. This expectation would stem from borderline Very High students being reclassified as High students, and thereby raising the average score in both the Very High (due to weaker students leaving) and High (due to stronger students entering) categories. This same expectation would hold for the Moderate vs. Very Low categories if students were in fact being reclassified between the two.

The fact that Figure 11 shows somewhat random trends in each category's student persistence rates during this time (most of which trend downward) indicates that students were being reclassified more dramatically than the minor shifts hypothesized above, and that the predictions were generally optimistic. The only category that trended noticeably upward was Very Low, indicating that students being reclassified into this category came from categories with higher persistence predictions (i.e. any other category). Further research that tracks individual student predictions across the semester would be necessary to confirm this conclusion.

Research Question 3: At What Rate do Different Demographic Groups at the College Persist?

As discussed in Chapter 2, numerous studies have demonstrated measurable differences in persistence rates among different demographic groups. Given this, it follows that considering demographics when analyzing persistence at the college is necessary. The goal in answering this question was to determine if similar differences existed at Mid-State Technical College, and if so, what areas of opportunity exist to improve persistence at the college.

The different demographics considered for this study included race (White, Hispanic, Asian, African American, and Native American), gender, course load (full-time vs. part time), and finance method (financial aid vs. self-paid). The study's findings are summarized below.

Race. As figure 12 shows, persistence varies widely when considering student race, but not necessarily in agreement with the results of the study cited in Chapter 2 (where Asian students persisted at a rate of 61.1%, compared to white students persisting at 56.9%, while Black (41.7%), Hispanic (41.7%), and Native American (35.8%) students persisted at significantly lower rates). Notably, both Hispanic (59.38%) and Native American (82.35%) students at Mid-State Technical College strongly outperformed the rates presented in the cited study, while White (70.67%), Asian (66.67%), and African American (44.44%) students at the college persisted at rates comparable to those found in the cited study.

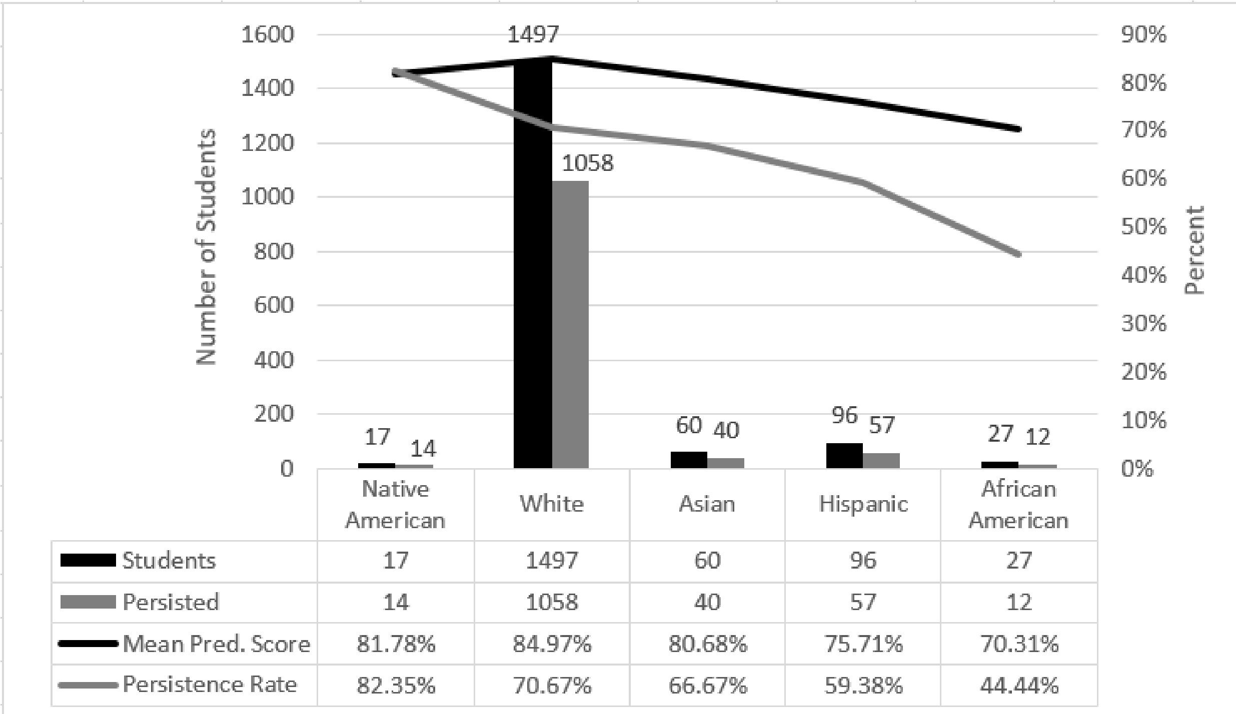


Figure 12. Persistence by race.

Of note in Figure 12 is the trend of the mean persistence score for each race, as it generally agreed with actual persistence. For instance, the three races that persisted at the highest rates (Native American, White, and Asian) all had mean persistence scores above 80%, while the two races that persisted at lower rates (Hispanic and African American) had lower mean persistence scores. This would appear to demonstrate the software's accuracy, as it correctly predicted persistence results in conflict with previous research. If the software's prediction algorithm considers student race, it appears that other factors influenced the predictions more heavily, yielding accurate results.

Whether the software considers race in its prediction algorithm was unclear, as top predictors were listed on a summary page in the software's interface only if the sample size was greater than one hundred; a condition only met by the White category. The small sample sizes among each race category should be noted. As discussed earlier, the software's accuracy for sample sizes under one hundred became erratic at times. While the results regarding race appeared to be accurate, this may be due in part to chance considering the samples size. If a second study produced similar results, one could more confidently conclude that the software accurately predicts persistence when considering student race.

Gender. Figure 13 shows persistence results when considering student gender. Of the 1,703 degree seeking students at the college, the 970 female students persisted at a higher rate (71.55%) than the 733 male students (67.39%). The mean persistence score for the female students at the college was higher (85.68%) than that of the male students (81.86%). Again, these mean persistence scores were higher than actual persistence rates, but the different between the two (3.82%) agreed with the difference in actual persistence rates (4.16%).

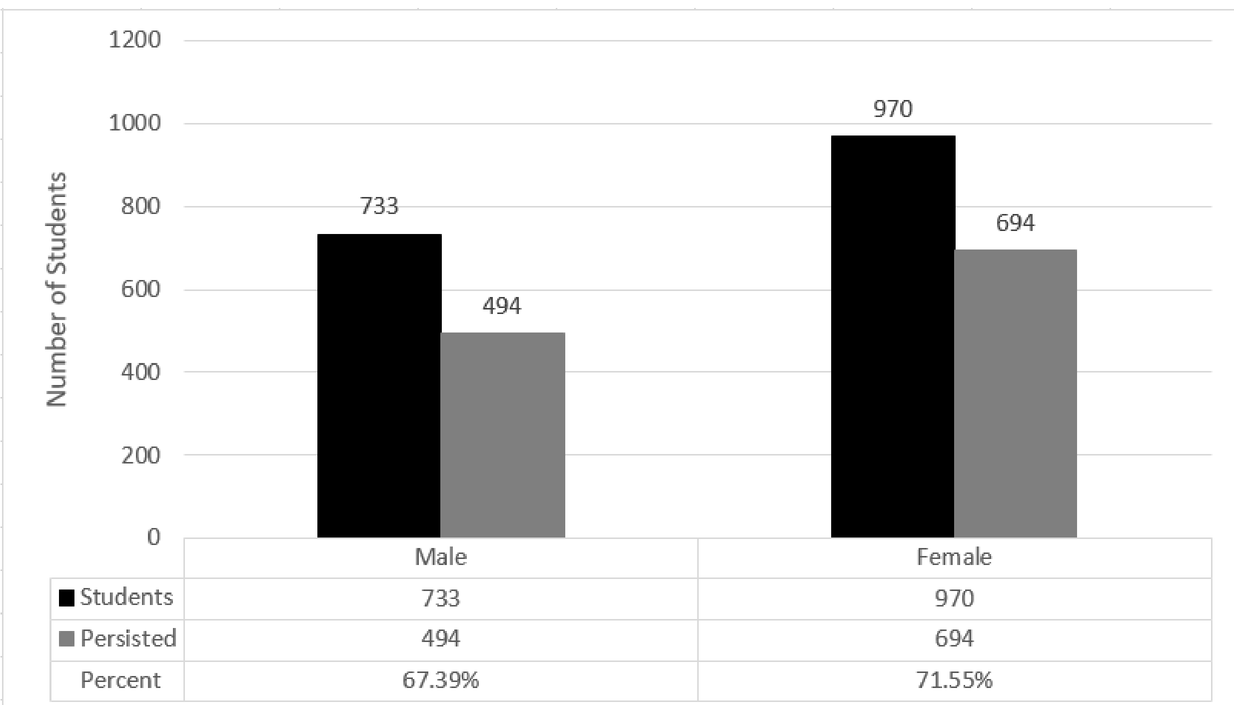


Figure 13. Persistence by gender.

Course load. When considering student course load, the 699 full-time students at the college persisted at a much higher rate (76.39%) than the 1,004 part-time students (65.14%).

Mean persistence prediction scores for the two groups were again higher than actual persistence rates, but the software scored the full-time students higher (87.47%) than it did the part-time students (80.52%). While the software predicted which group would persist at a higher rate, the difference in prediction scores between the two groups (6.95%) was considerably less than the difference in actual persistence rates between the two groups (11.25%), indicating a level of inaccuracy.

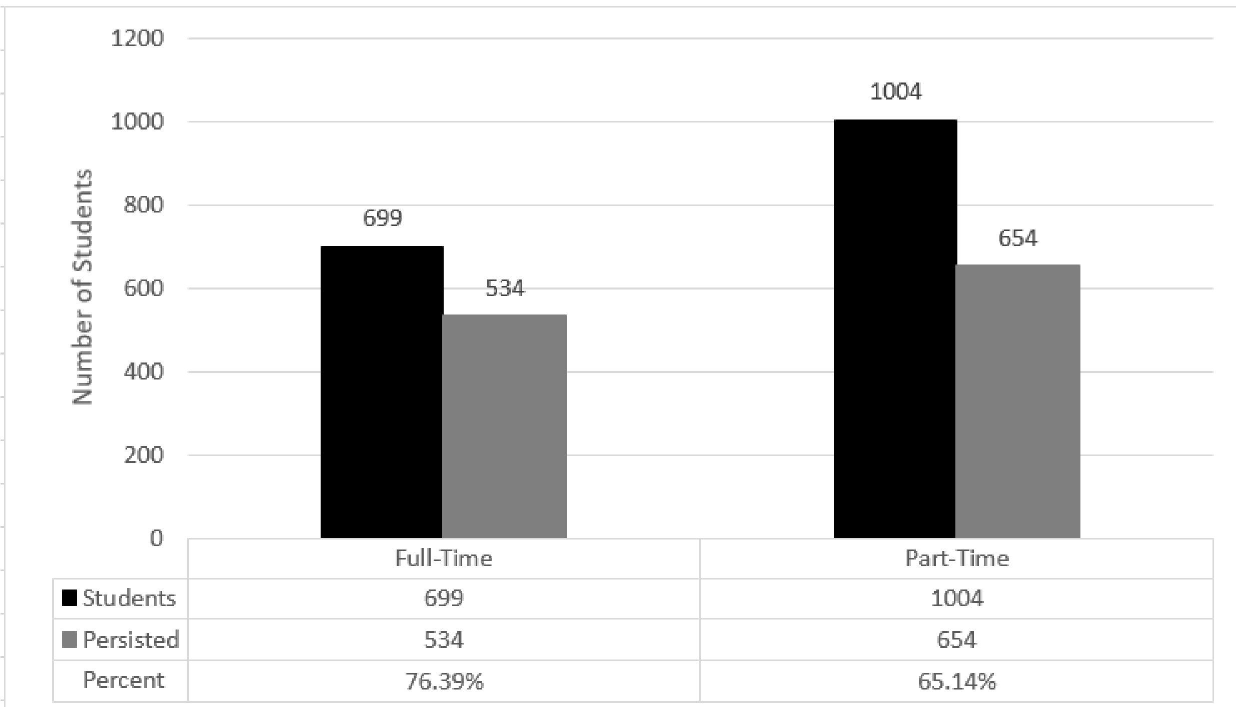


Figure 14. Persistence by course load.

Finance method. The final demographic considered in the study was the finance method of the student: financial aid vs. self-paid. As shown in Figure 15, nearly two-thirds of the students at the college receive some form of financial aid. Those receiving financial aid persisted at a much higher rate (77.64%) than those who did not (64.07%).

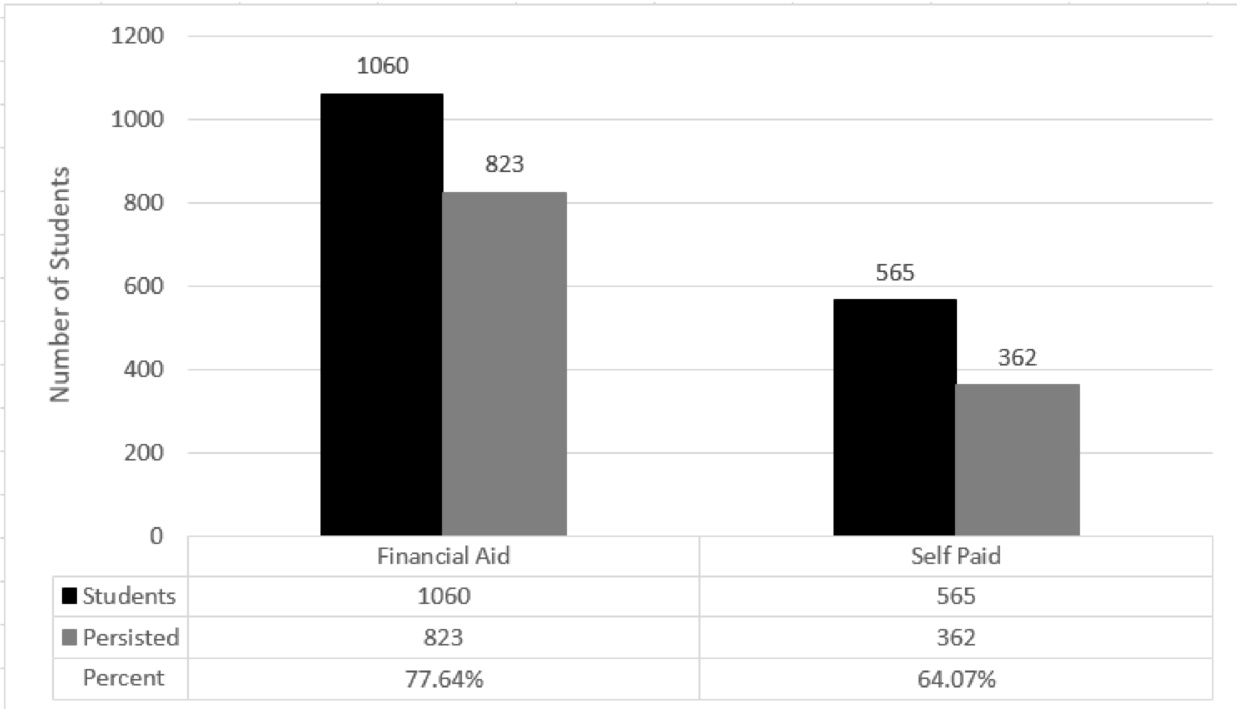


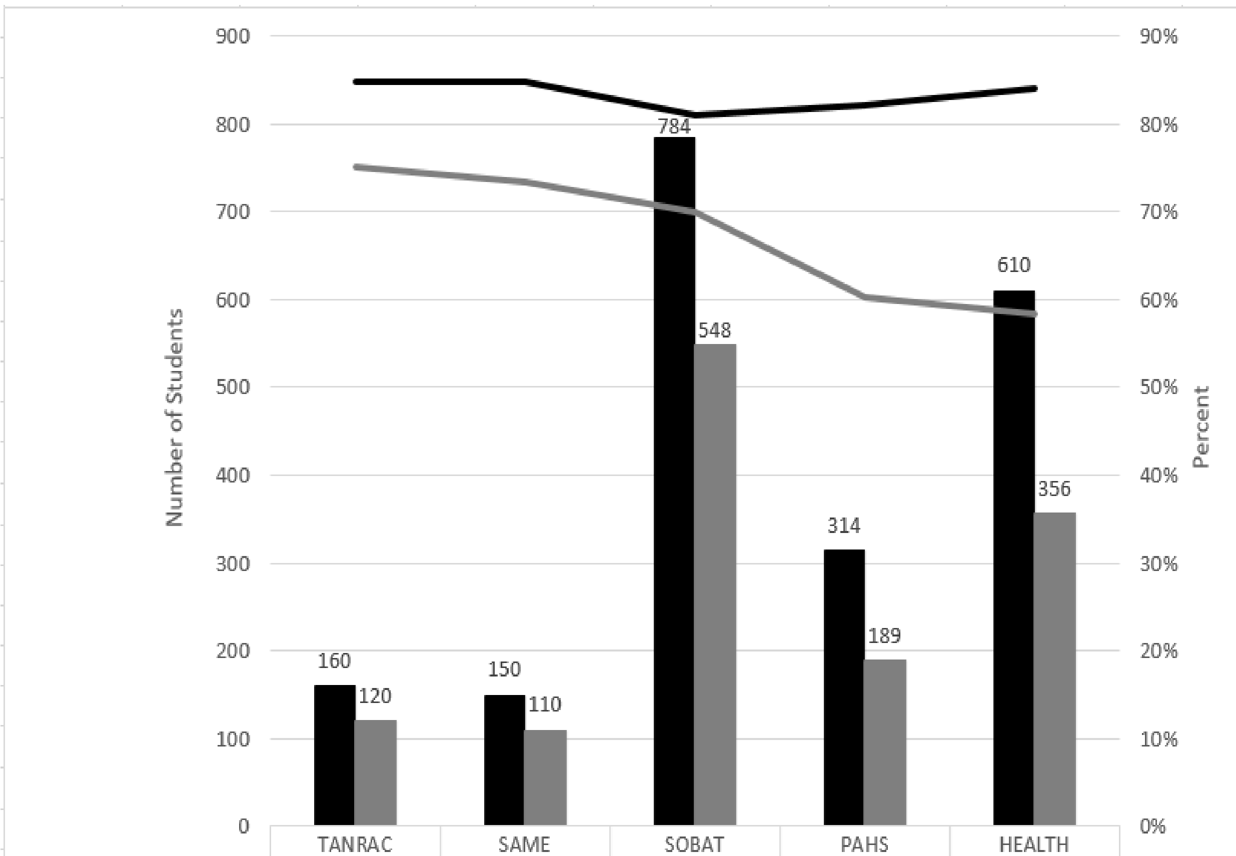
Figure 15. Persistence by finance method.

Assuming that those not receiving financial aid paid for their education themselves (this may not always be true, but is difficult to discern), this result is somewhat surprising. One might think that those paying for their own education would be more determined to persist than those having all or a portion of it paid for them. The likely explanation for this is that many forms of financial aid require students to be full-time, and as discussed above, full-time students persist at much higher rates than part-time students.

Regarding the software's predictions, the mean prediction scores for these two groups agreed with this result, but did not vary by nearly as much as the actual persistence rates (84.81% for financial aid students vs. 82.62% for self-paid students). Again, this seems to indicate inaccuracy in the software's predictions.

Research Question 4: At What Rate do Different Schools Within the College Persist?

The programs offered at Mid-State are organized into six "schools." These include the School of Transportation, Agriculture, Natural Resources, and Construction (TANRAC), School of Advanced Manufacturing and Engineering (SAME), School of Business and Technology (SOBAT), School of Protective and Human Services (PAHS), School of Health (HEALTH), and School of General Education and Learning Resources (SOGEALR).



	TANRAC	SAME	SOBAT	PAHS	HEALTH
Students	160	150	784	314	610
Persisted	120	110	548	189	356
Mean Pred. Score	84.88%	84.83%	81.05%	82.18%	84.09%
Persistence Rate	75.00%	73.33%	69.90%	60.19%	58.36%

Figure 16. Persistence by school.

Figure 16 shows the enrollment, persistence, and prediction scores for each school except the School of General Education and Learning Resources, in which only a handful of students pursued a degree. That is not to say that a large number of students were not enrolled in general education courses required for their chosen degree, but rather, that only a few pursued degrees specifically in general education. The filters used in the software to generate the data for each school used the students' chosen degree to categorize them into schools, and therefore, only yielded six students in the SOGEALR. Given such a small sample size, this school was omitted from the analysis.

Among the remaining five schools, enrollment and persistence varied widely. TANRAC and SAME both had fewer students enrolled than the other schools (160 and 150 respectively), but these students persist at a higher rate (75.00% and 73.33% respectively). SOBAT had the highest enrollment at the college (784), and a persistence rate (69.90%) comparable to TANRAC and SAME. PAHS and HEALTH both had moderate to high enrollments (314 and 610 respectively), but students persisted in these schools at a rate of 60% or less.

The reasons for these results are unclear, and might include some of those discussed in Chapters 1 and 2, such as class size, selectivity, and rigor. Further research would be necessary to determine what is causing student persistence to vary among the different schools at the college.

A final item of note regarding Figure 16 is that, despite actual persistence varying widely among the different schools, the prediction scores generated for each school did not. All schools had mean prediction scores above 80%, with HEALTH among the highest (84.09%). When compared to actual persistence, it appears that the software's predictions were not accurate across all schools.

Chapter V: Discussion, Conclusion and Recommendation

The purpose of this study was to determine whether the newly acquired Illume Students software predicts student persistence with reasonable accuracy. Further, it intended to determine how persistence at the college varied relative to time, demographic, and school. The findings of the study are summarized below.

Discussion

The above study considered data for the Spring 2018 semester, with persistence into the Fall 2018 semester. Persistence rates from Fall to Spring semester are likely different, and should be analyzed in a separate study that captures data during the Fall and compares it to enrollment data for the ensuing Spring. This study could be compared with Spring to Fall studies, with the expectation that persistence rates will differ between them, while other trends, such as critical times during the semester, may not.

Conclusions

The study determined that the software's predictions were reasonably accurate when considering student groups of adequate sample size (100 or more students, with at least 15-20 in each prediction category). When these two conditions were not met, however, the accuracy of the predictions often became erroneous. This was most apparent when considering the different races of students at the college, as none of the minority groups met the two sample size criteria explained above.

This limitation is concerning, as most users of the software (namely faculty and counselors) intend to identify at risk students in hopes of offering them timely guidance that will improve their outcomes. If the predictions are not reasonably accurate on an *individual* level, the software's value in identifying *individuals* at risk becomes questionable. Couple that with the two week delay in prediction response found in the study, and the information necessary to provide *timely* guidance to at risk students may arrive too late to be useful.

That being said, the software's accuracy for large groups of students may be helpful at an administrative level, and this accuracy provides hope that the algorithm can be improved to yield more accurate and timely predictions at the individual level. To that end, faculty now have the ability to keep attendance in Blackboard, which is one of the databases that the software uses to make its predictions. Assuming the prediction algorithm considers this new information, the timeliness and accuracy of its predictions is expected to improve. If given proper consideration, a student's absence from class should cause his prediction score to decrease accordingly, and in nearly real-time (i.e. within one or two days, not two weeks).

Regarding how persistence at the college varies over time, it is apparent that the first three to five weeks of the semester are critical, with many of those that do not persist deciding to do so very early in the semester in the form of a no-show student or one that withdraws from classes prior to the refund deadline.

Similarly, the enrollment date for the next term is a critical time to promote persistence. The students' intentions to persist (or not) are revealed by their actions during this time. If they enroll in the following term on or near this date, it is almost a certainty that they will persist, both because they have enrolled, and because they are motivated/organized enough to enroll at such an early date. Interestingly, one of the "powerful predictors" the software often uses to make its predictions is how far in advance a student enrolls for classes.

Conversely, students that do not enroll in a timely fashion demonstrate that they do not intend to persist at the college, are unsure, disorganized, unmotivated, or have a hold on their account (financial or academic). In any case, this action indicates that the person is less likely to persist than someone who does not demonstrate these qualities.

A third critical time for student persistence is near the end of the semester. During this time, the prediction scores of many students fluctuated enough to put them into a different prediction category.

It is worth noting that prediction scores did not vary noticeably near the week of spring break. One might expect that some struggling students might not return to class after the week off, but this was not apparent in the data. If this phenomenon was occurring, it went unnoticed by the software.

When considering different demographic groups at the college, the most apparent indicator of persistence was the course load of the student. Full-time students persisted at a much higher rate than part-time students. This was evident in the financial aid group as well: most students receiving financial aid are required to have a full-time course load. To a lesser degree, female students persisted at higher rates than males.

Regarding the different schools at the college, enrollment and persistence varied widely, while the prediction scores for these groups remained rather constant. This indicated that the software's predictions were somewhat inaccurate with these filters applied.

Recommendations

This study has provided insight into the accuracy of the persistence predictions made by the newly acquired Illume Students software at Mid-State Technical College. Given its accuracy for large groups of students, it could be a useful tool for administrators to estimate persistence rates at the college. Applied to smaller groups, its current value is suspect, and strict reliance on the software's predictions to deploy improvement efforts to individual students could be misguided.

That being said, as more data is made available for the software to make its predictions, such as real-time student attendance, prediction accuracy may improve to a usable level. Repeating this study considering future Spring to Fall student data would help determine if this is the case. If the study were to be repeated, it would be beneficial to have it begin the first week of the semester (rather than the third) to better analyze the critical early weeks as they relate to student persistence and prediction accuracy.

Considering other student groups with small sample sizes might provide a better understanding of how the prediction accuracy varies with sample size, and if accuracy at the individual student level has been achieved. Until it is shown that this level of accuracy is present, faculty and counselors should use the software's predictions with caution, relying more heavily on the relationships developed with students and the open communication fostered therein.

References

Berger, J. & Braxton, J. (1998). Revising Tinto's interactionalist theory of student departure through theory elaboration: Examining the role of organizational attributes in the persistence process. *Research in Higher Education, 39*(2)*,* 103-119.

Civitas Learning (2018). *Illume.* Retrieved from https://www.civitaslearning.com

Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention, 13*(1), 17-35.

Lambeth, J., Joerger, R., & Elliot, J. (2009). Implications for focusing research in career and technical education and workforce development. *Career and Technical Education Research, 34*(3), 137-153.

Maas, J. (2008). *Is the severity of mental health needs and the degree of mental health services received predictive of student retention at UW-Stout* (Unpublished master's thesis). University of Wisconsin-Stout, Menomonie, WI.

Mid-State Technical College. (2018a). *Additional information.* Retrieved from http://www.mstc.edu/student-resources/policies/additional

Mid-State Technical College. (2018b). *Payment & refund policies.* Retrieved from http://www.mstc.edu/student-resources/registration/payment-refund

Milem, J., & Berger, J. (1997). A modified model of college student persistence: Exploring the relationship between Astin's theory of involvement and Tinto's theory of student departure. *Journal of College Student Development*, *11,* 386-400.

National Center for Educational Statistics (2017, April). *Undergraduate retention and graduation rates.* Retrieved from https://nces.ed.gov/programs/coe/indicator_ctr.asp

National Student Clearinghouse Research Center (2017, June 12*). Snapshot report – first-year persistence and retention.* Retrieved from https://nscresearchcenter.org/snapshotreport28-first-year-persistence-and-retention/

Peltier, G., Landen, R., & Matranga, M. (2000, February 1). Student persistence in college: A review of research. *Journal of College Student Retention: Research, Theory, & Practice, 1*(4), 357-375.

Schmidt, H. G., Cohen-Schotanus, J., van der Molen, H. T., Splinter, T. A. W., Buldte, J., Holdrinet, R., & van Rossum, H. J. M. (2009). Learning more by being taught less: A “time-for-self-study” theory explaining curricular effects on graduation rate and study duration. *Higher Education, 60*(3), 287-300.

Seidman, A. (2005, March). Minority student retention: Resources for practitioners. *New Directions For Institutional Research, 2005*(125), 7-22.

Strahota, H. (2016, June 21). *Average student has better chance (77%) of graduating at selective universities compared to open access schools (51%), Georgetown University analysis finds.* Retrieved from https://cew.georgetown.edu/wp-content/uploads/MismatchFisherUT_pressrelease_6-21-16.pdf

Sullivan, P. (2008, July). Measuring “success” at open admissions institutions: Thinking carefully about this complex question. *College English, 70*(6), 618-632.

Tinto, V. (1999). *Taking retention seriously: Rethinking the first year of college.* Chicago, IL: University of Chicago Press.

Tinto, V. (2012a). Enhancing student success: Taking the classroom success seriously. *The International Journal of the First Year in Higher Education, 3*(1)*,* 1-8.

Tinto, V. (2012b). Leaving college: Rethinking the causes and cures of student attrition. *NACADA Journal, 19*(2), 5-9.

Tinto, V. (2017). Reflections on student persistence. *Student Success, 8*(2), 1-8. doi: 10.5204/ssj.v8i2.376

Wright, S., Jenkins-Guarnieri, M., & Murdock, J. (2012). Career development among first-year college students: College self-efficacy, student persistence, and academic success. *Journal of Career Development, 40*(4). 292-310.