Research Brief: Predicting 1-Year Retention – A Decision Tree Model Approach

UNIVERSITY OF CALIFORNIA MERCED

March 2022

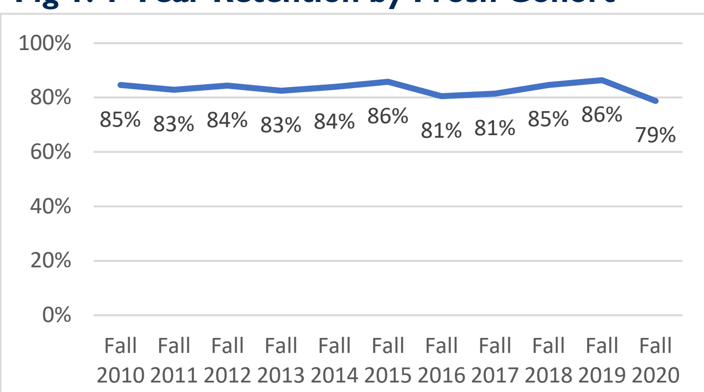
Cinnamon L. Danube & Corinne Townsend

This brief describes the results of an analysis conducted using a decision tree methodology to answer the question: What are the best opportunities for improving the 1-year retention rate for incoming frosh at UC Merced? Data from four frosh cohorts were combined, and 43 demographic, academic/start of term, and end of first term variables were considered across nine possible models. In prioritizing model accuracy and parsimony, the key predictors of 1-year retention were grades in the gateway courses of Writing 001 (Academic Writing) and Math 005 (Pre-Calculus), and to a lesser extent Writing 010 (College Reading and Composition), in the first term. Other variables did appear in the chosen model but were either not actionable or were not useful given the small impact they would have on retention at the campus level. The results are promising in that they suggest multiple routes for improving 1-year retention rates at the campus level by bolstering existing efforts or developing new approaches to improving new student success in these gateway courses.

Background

Despite efforts to improve, the 1-year retention rate for new frosh has remained consistent over time ranging from 79% (during the Covid-19 pandemic) to 86% (Fig 1). This indicates that new approaches are needed and provides an opportunity to consider which might be the most effective for improving student success. The goal of this analysis was to use a machine learning approach called a decision tree model to identify impactful opportunities for improving the 1-year retention rate at UC Merced. Our current target for 1-year retention is 90%, leaving, on average, about a 7% opportunity for improvement. A good rule of thumb given present undergraduate enrollments is that retaining an additional 140 students would improve 1-year retention by the desired 7% (i.e., 20 students for every 1% improvement).

Fig 1. 1-Year Retention by Frosh Cohort



Sample and Methodology

We combined data from four fall frosh cohorts – 2015 to 2018 – for our analysis (n = 8348). We intentionally did not include cohorts during the Covid-19 pandemic due to the disruptions to the student experience during that time.

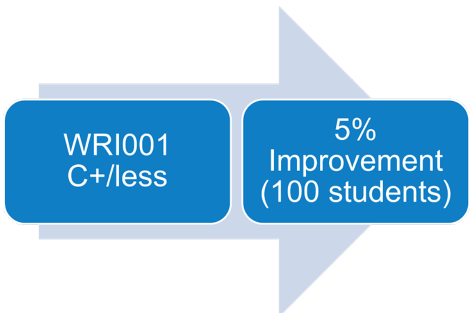
We considered 43 variables (see Appendix A) across three categories: demographic (e.g., race/ethnicity), academic/start of term (e.g., high school GPA, number of credits enrolled), and end of first term (e.g., grades in gateway courses). A strength of the decision tree approach is the ability to include many variables with minimal coding. However, unlike other models (e.g., regression), there are no statistical controls – each variable is considered independently.

The analysis was conducted in SPSS Modeler and used a CHAID decision tree model to predict student enrollment at the start of the third term (i.e., retention to 1-year). Overall model accuracy was explored across training (67% of sample) and testing (33% of sample) sets. Nine models were explored (see Appendix B), which varied characteristics such as stopping rules, costs, and tree depth. In selecting the final model, we prioritized model accuracy and parsimony. The final model (see Appendix C) was 75% accurate overall in the training set and 76% accurate in the testing set. Critically, it was effective at predicting both which students would be enrolled (78% recall/sensitivity) and <u>not</u> enrolled (62% specificity) at 1-year.

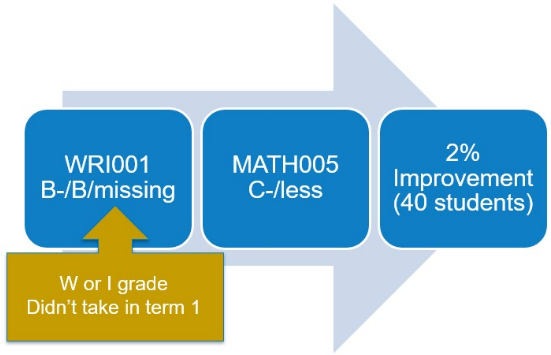
Targets for Improving 1-Year Retention

We identified five key branches for improving 1-year retention at the campus level, which indicated that improving student performance in the gateway courses of WRI001 (Academic Writing), MATH005 (Pre-Calculus), and to some extent WRI010 (College Reading and Composition) would be the most effective for improving overall retention to 1-year.

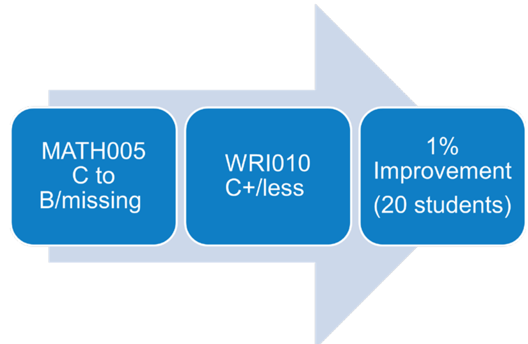
Key Branch #1: WRI001. This branch conveys the importance of WRI001 grades in the first term for retention. <u>If we were able to retain all students who received a poor grade in WRI001 (C+ or less) in the first term and who subsequently left the university, we could improve the 1-year retention rate by 5%.</u> This is roughly equivalent to retaining an additional 100 students.



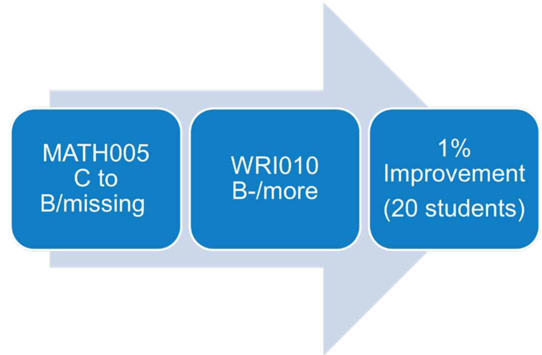
Key Branch #2: WRI001 & MATH005. This branch conveys the importance of both WRI001 and MATH005 grades in the first term for retention. The students in this branch have two characteristics – they received a WRI001 grade of B-, a grade of B, or had a missing grade <u>and</u> received a MATH005 grade of C- or less. Importantly, students can have a missing grade at the end of the first term for several reasons: a) they withdrew from (W) or received an incomplete (I) in the course (relatively rare), b) they tested out of the course and did not need to take it, or c) they tested into the course but did not take it in their first term. If we were able to retain all students with these characteristics in the first term who subsequently left the university, we could improve the 1-year retention rate by 2% - roughly equivalent to retaining an additional 40 students.



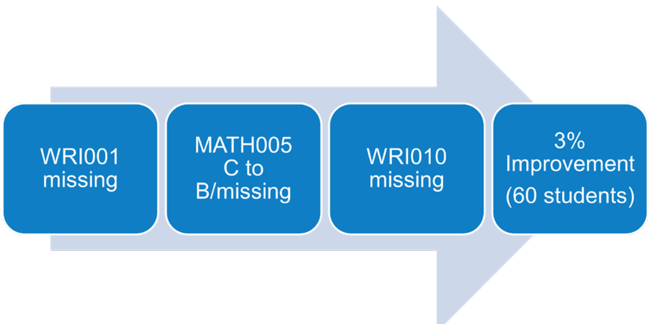
Key Branch #3: MATH005 & WRI010. This branch conveys the importance of MATH005 and WRI010 grades in the first term for retention. The students in this branch have two characteristics – a MATH005 grade between C to B or a missing grade <u>and</u> a WRI010 grade of C+ or less. If we were able to retain all students with these characteristics in the first term who subsequently left the university, we could improve the 1-year retention rate by 1% - roughly equivalent to retaining an additional 20 students.



Key Branch #4: MATH005 & WRI010. This branch again conveys the importance of MATH005 grades, but when considered along with branch #3 indicates that WRI010 grades have an unclear association with retention. The students in this branch have two characteristics – a MATH005 grade between C to B or a missing grade (same as branch #3) <u>and</u> a WRI010 grade of B- or more. Because branch 3 and 4 are identical excepting that they each indicate a different subset of WRI010 grades, MATH005 grades appear to be the most critical for retention. That aside, if we were able to retain all students with these characteristics in the first term who subsequently left the university, we could improve the 1-year retention rate by 1% by retaining about 20 additional students. Combining branches 3 and 4 to focus on MATH005 grades would yield a 2% (40 student) improvement.



Key Branch #5: No WRI & MATH005. This branch again conveys the importance of MATH005 grades, but also suggests that missing grades in WRI001 and WRI010 are important for retention. The students in this branch have three characteristics – a missing grade in WRI001, a MATH005 grade between C to B or a missing grade (same as branch #4), and a missing WRI010 grade. If we were able to retain the students with these characteristics in the first term who subsequently left the university, we could improve the 1-year retention rate by 3% - roughly equivalent to retaining an additional 60 students.



Non-Useful or Actionable Variables. A reflection on the variables that the decision tree indicated were not important is also useful for thinking about retention at the campus level. Demographic variables (e.g., gender and race/ethnicity) were not critically important at the campus level for improving one year retention. Though the campus does have equity gaps to address, they do not appear to be critical for 1-year retention at the campus level. Similarly, academic preparation variables (e.g., high school GPA) and start of term variables (e.g., number of credits enrolled, living on campus) were not in the final model and so do not appear critically important for improving 1-year retention.

Most simply put, this indicates that how students perform academically in their first term gateway courses is more important than pre-matriculation and start of term factors for improving 1-year retention. From an intervention standpoint, this indicates the campus has a great potential for improvement by implementing programs and creating opportunities to help our students be successful in their gateway courses.

Possible Strategies for Improving 1-Year Retention

UC Merced already employs strategies for improving 1-year retention rates, which can be informed by the results of this analysis. For example, it may be useful to think more about how we reach out to students to inform them about available resources and about whether we are reaching the students who need help most. Can we improve student use and efficacy of learning support (Writing Center, Math Center, STEM Resource Center, tutoring, Bright Success Center)? What pedagogical approaches could be improved in gateway courses and can we provide other direct support in courses? As the co-curricular experience can also be critical for course performance, what opportunities are there for improving student engagement and support outside of class? For example, can we leverage campus Living Learning Communities to improve student success in their gateway courses?

The campus could also consider employing new strategies for improving 1-year retention. For example, we can think about the placement tests used to determine which students must take WRI001 and MATH005 and whether those tests are effective. Have we identified appropriate cutoff scores, are there ways for students to engage in test prep to ensure the most accurate scores, are there other ways for students to test out of courses like AP credit, can we partner community colleges or summer bridge so students who need these courses can take them before their first fall semester? We could also consider working with K-12 partners to ensure students receive the preparation they need to be successful at UC Merced. With regards to the way we support students in these courses, like community colleges, we could think about employing co-requisite models where students take the college level course (e.g., calculus – MATH011 or MATH021) along with elements intended to refresh the knowledge that may be lacking (e.g., pre-calculus, which is MATH005). We could also think about adding more supplemental instruction or improving other aspects of our pedagogy in gateway courses (e.g., flipped classrooms and other types of experiential/applied learning). Finally, we could use student data to create models that will help us identify which students might struggle in gateway courses (e.g., CatCourses alerts, predicted probability of passing) so that we can offer those who might struggle additional support (e.g., academic coaching).

Contact Us

University of California, Merced
Center for Institutional Effectiveness
Institutional Research and Decision Support
CIE.UCMERCED.EDU

Appendix A: Complete Model Variables List

Demographic	Academic/Start of Term	End of Term
• Pell eligibility status • First generation status • Region of high school • Gender • Race/ethnicity • Residency status	• High school GPA • Admit type • A-G & honors courses • UC's admitted to • School & major • Housing status • Credit hours • Summer units • Gateway courses enrolled (CHEM001, MATH005, MATH011, MATH021, WRI001, WRI010, BIO001)	• Gateway course grades earned (CHEM001, MATH005, MATH011, MATH021, WRI001, WRI010, BIO001)

Appendix B: Decision Tree Model Statistics

Statistic	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5	Tree 6	Tree 7	Tree 8*	Tree 9
Testing Set Accuracy	85%	85%	85%	85%	85%	85%	74%	76%	75%
Training Set Accuracy	85%	85%	84%	85%	84%	85%	74%	75%	75%
Testing Set Recall/Sensitivity (Enrolled)	98%	98%	98%	98%	98%	97%	75%	78%	78%
Testing Set Specificity (Not Enrolled)	22%	22%	16%	22%	16%	26%	64%	62%	62%
Number of Variables in Tree	9	11	9	7	5	6	11	7	8

*Selected tree

Appendix C: Final Decision Tree Model