

High-Level Targets to Improve Six-Year Graduation Rates for Pell & URM First-Time Freshmen: Iterative Decision Tree Modeling Results

Contributors: Laurie McConville, Peter Owens, Ian Donatello, Tom Rosenow

Chico State Office of Institutional Research

September 8, 2021

Executive Summary

Institutional Research staff performed quantitative modeling using two decision tree algorithms to identify academic and demographic characteristics that influence six-year graduation outcomes for Pell-receiving and Underrepresented Minority (URM) first-time freshmen (FTF) in the Fall 2012 – Fall 2014 cohorts. Six-year graduation outcomes for individual students were modeled using two “decision tree” model types. The results of these models were supplemented by the identification of factors that influence four-year graduation and one- and two-year retention, with the same inputs used across these model sets; the FTF cohorts used in these supplemental analyses are the Fall 2012 – 2016 and Fall 2012 – 2017 cohorts, respectively.

*Results indicate that **First-term GPAs for Pell-receiving and URM FTF above an initial threshold of about 1.8 were needed to make graduation within six years attainable for most students (21% graduation rate below vs. 68-69% graduation rate above), while GPAs at or above a secondary threshold of about 2.9 made graduation within six years very likely (78-79%) to occur.** Supporting models further indicate that Pell and URM FTF with **Entry GPAs (i.e., high school GPAs) below 3.2 (for URM) to 3.4 (for Pell), and who either have undeclared majors, or who have declared majors in Humanities & Fine Arts, Natural Sciences, or Engineering, Computer Science, & Construction Management (for URM), or in Humanities & Fine Arts, Natural Sciences, Engineering, Computer Science, & Construction Management, or Business (for Pell), have a decreased likelihood of graduating within six years.** These latter two criteria (high school GPA and major) are generally known upon matriculation. They can therefore be used to coordinate policy changes and services early in a student’s first term. Further modeling identified **important differences in the likelihood of six-year graduation by first-generation status and gender for Pell-receiving FTF, and by gender and first-term attempted units for URM FTF, which provide further potential criteria for coordination of policy and student services.***

Modeling is ongoing and iterative, with all results subject to future revision by IR staff. All model-based predictions are made using probabilistic, not deterministic, criteria and are therefore subject to variability. As models are further refined, and as the campus makes changes to underlying social and academic conditions, further variability in rankings of indicator importance will inevitably occur. **Due to limited data availability, model inputs largely represent the academic, and not behavioral or social, characteristics of students, and results are therefore likely to be affected by the omission of confounding factors that influence both model inputs (e.g., First-term GPA) and outcomes (e.g., six-year graduation).** It is critically important for the Chico State community to consider means by which to collect these types of data more thoroughly at a systematic level, both through existing instruments as well as by creating new, empirically-supported means for their collection at the college

and departmental levels. IR staff will continue to work with campus partners to update model inputs and results as systematic social and behavioral data become more fully available.

Data and Methodology

The principal analysis of six-year graduation outcomes utilized institutional enrollment data from the CSU, Chico Enrollment Reporting System (ERS) for First-Time Freshmen students in the Fall 2012 – Fall 2014 cohorts. Table 1 shows counts and proportions for first-time freshmen in the three cohorts by gender, first-generation status, underrepresented minority (URM) status, and whether or not students received Pell Grant support. As can be seen, female students made up a larger proportion of each cohort than males. We can also note that a high proportion of URM students and Pell-receiving students in each cohort were also the first in their families to attend a four-year postsecondary institution.

Table 1: Selected Demographic Characteristics of First-Time Freshmen Cohorts, Fall 2012 – Fall 2014

Cohort Term	Gender	First-Generation Status	Total Headcount	URM Headcount	% URM	Pell Headcount	% Pell
Fall 2012	Female	First-Generation	379	236	62.3%	301	79.4%
		Not First-Generation	1,057	238	22.5%	303	28.7%
		Subtotal	1,475	482	32.7%	608	41.2%
	Male	First-Generation	259	162	62.5%	208	80.3%
		Not First-Generation	912	177	19.4%	181	19.8%
		Subtotal	1,239	343	27.7%	393	31.7%
Subtotal			2,714	825	30.4%	1,001	36.9%
Fall 2013	Female	First-Generation	373	253	67.8%	299	80.2%
		Not First-Generation	927	243	26.2%	296	31.9%
		Subtotal	1,328	501	37.7%	602	45.3%
	Male	First-Generation	226	162	71.7%	194	85.8%
		Not First-Generation	739	173	23.4%	189	25.6%
		Subtotal	1,012	342	33.8%	390	38.5%
Subtotal			2,340	843	36.0%	992	42.4%
Fall 2014	Female	First-Generation	541	397	73.4%	448	82.8%
		Not First-Generation	1,065	306	28.7%	361	33.9%
		Subtotal	1,661	720	43.3%	823	49.5%
	Male	First-Generation	344	229	66.6%	274	79.7%
		Not First-Generation	885	242	27.3%	243	27.5%
		Subtotal	1,284	478	37.2%	529	41.2%
Subtotal			2,945	1,198	40.7%	1,352	45.9%
Grand Total			7,999	2,866	35.8%	3,345	41.8%

Besides baseline demographic characteristics, the input variables used to model six-year graduation outcomes focused on student academic characteristics during the first year of study, particularly the first term. The inputs currently used in each model are as follows:

California State University Chico

- Hispanic / Latino status
- Concentration / Major
- College of Major
- Department of Major
- Enrollment County
- Residence Area
- Entry Grade Group
- Entry GPA
- One-Term Continuation
- First-term GPA
- First-term Units Attempted
- First-term Units Passed
- IPEDS Race / Ethnicity
- Gender
- URM Status
- Pell Status
- Age Group at Census
- EOP Status
- Parent's Education
- GE Math Status
- GE Critical Thinking Status
- A4 Milestone Level
- A4 Milestone Completion
- A4 Milestone Completion Type
- A2 Milestone Level
- A2 Milestone Completion
- A2 Milestone Completion Type
- A2 Milestone Level

It should be noted that each model is limited by the types of institutional data that are currently available on a systematic and historical basis to Institutional Research. For example, the models largely use academic inputs to predict academic outcomes: First-term GPA is used as a predictor of graduation / non-graduation, but First-term GPA itself is *also an academic outcome* that is influenced by other prior factors. Furthermore, the lack of systematic institutional data on key social and behavioral topics such as socioeconomic status, engagement with student services, organizational affiliations, or student social experiences, means that the models cannot account for the effects of these likely-influential, but unfortunately omitted, factors. Institutional Research will continue to work with its campus partners to update and refine these and other models as new data on such important topics become available.

Six-year graduation outcomes for individual students were modeled using two “decision tree” algorithm models: Classification and Regression Trees (CRT) and Chi-Square Automatic Interaction Detection (CHAID). All modeling was performed using SPSS Data Modeler 18.2.1. Models using each of the two decision tree algorithms utilized data for three subgroups of first-time freshmen: non-Pell/URM students, URM students, and Pell-receiving students. Because initial models were largely dominated by First-term GPA and units attempted / passed, each of the two models for the three groups of students were also performed with the First-term GPA / units attempted / units passed inputs excluded, in order to identify secondary factors that were significant to graduation outcomes. This yielded 12 total analytic models for the study. The results of these models were supplemented by the identification of factors that influence four-year graduation and one- and two-year retention, with the same inputs used across

these model sets; the FTF cohorts used in these supplemental analyses are the Fall 2012 – 2016 and Fall 2012 – 2017 cohorts, respectively.

Visually, decision trees can be represented as a flowchart-like structure where “parent” nodes are increasingly split into two or more “children” nodes. Each parent-children node relationship is based on the model’s identification of input variable thresholds that provide the greatest possible leverage in predicting the outcome (I.E., graduating within six years), and can be based on nominal, ordinal, and interval-ratio data types. These inputs are ranked hierarchically in the decision tree, with the most impactful input values close to the base of the tree and less-impactful predictor values further away from the base. Because of this hierarchical ranking of inputs based on predictive value, students sorted into initial children nodes will tend to exhibit larger differences in the predicted probability of achieving the outcome than students in nodes that are further away from the base.

An example of a decision tree produced using CRT is shown in Figure 1. In the parent node (Node 0), the baseline predicted probability of graduating within six years (represented as “1.000” in the node) is 71%. In the two children nodes (Nodes 1 and 2), we can see that the CRT model has sorted this initial group into two subgroups, each with a very different predicted probability of graduation, based on an identified threshold value: whether a student had a First-term GPA that was greater than or less than 2.13. For students who were above the threshold, their predicted probability of graduating in six years was about 80%; students below the threshold, in contrast, were predicted to have about a 30% chance of graduating in six years. This initial factor, therefore, maximizes the differences in chances of graduation between the two nodes, based on the available inputs in the model. Node 1 has no children nodes, meaning that the CRT model did not identify any further thresholds that might explain any differences in the chance of graduation for these 265 students. Node 2, which has 1,263 students and is therefore likely to have more variation in student characteristics, does have two children nodes based on a second identified First-term GPA threshold: 3.05. Students above this new threshold have an even better-predicted probability of graduation in six years (85%), while those with a GPA between 2.13 and 3.05 have a slightly lower predicted probability (74%).

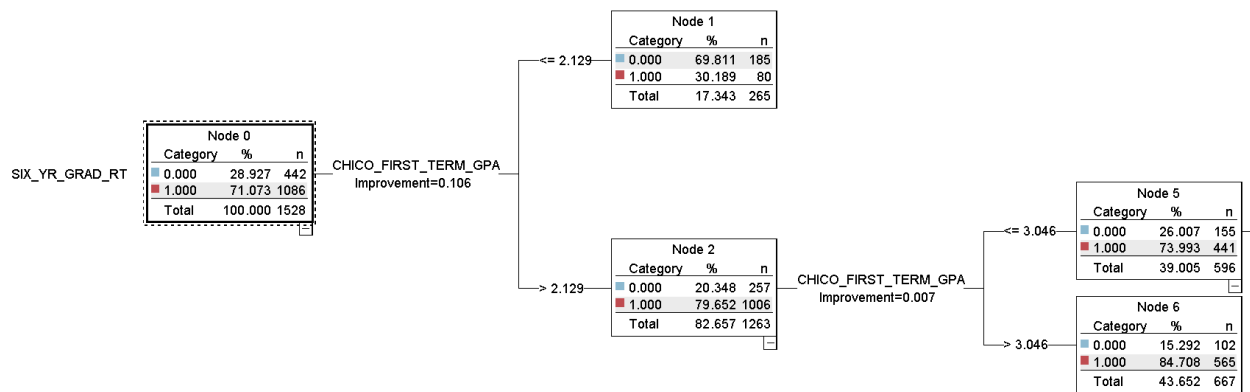


Figure 1. A partial CRT Decision Tree Predicting Six-Year Graduation for non-Pell / non-URM First-Time Freshmen

It should be noted that the decision tree models are inherently probabilistic in nature: based on the variability of the input variables used, and the relation of these characteristics to the outcome, each student in the data was sorted by the models into an increasingly complex network of either / or “bins” (E.G., either above or below a given GPA threshold), and each of these bins was assigned a predicted

probability for the outcome to occur. Despite their probabilistic nature, the models are able to correct for inevitable errors in this classification and sorting process. Specific weights were therefore applied so that each model would be less likely to misclassify students as having graduated within six years, when they in fact, did not. These misclassification weights were applied because the inaccurate prediction of graduation is especially costly, relative to the inaccurate prediction of non-graduation – the latter students still graduated, while the former did not. This is based on the rationale that it is far better to over-serve students who are likely to graduate anyway than to under-serve students who are unlikely to graduate.

Results

First-Term GPA

Of all first-year academic inputs noted above, First-term GPA was consistently the strongest predictor of FTF graduation within six years. Both CRT and CHAID models for Pell and URM students consistently identified initial First-term GPA thresholds, *below which* graduation within six years was extremely unlikely. CRT models also identified secondary thresholds, after the decision tree was split by the initial First-term GPA threshold, *above which* graduation within six years became much more likely. The initial First-term GPA thresholds, as the single most important value to sort students by their likelihood of graduation, can therefore be operationally understood as *working to avoid a bad outcome*, whereas the secondary First-term GPA thresholds can be understood as *working to make a good outcome more likely*.

- URM FTF First-term GPA Thresholds
 - Initial: > 1.84
 - Secondary: > 2.89
- Pell FTF First-term GPA Thresholds
 - Initial: > 1.82
 - Secondary: > 2.94

While the prediction of these threshold levels in the CRT and CHAID models are based on probabilistic criteria, we can also apply these thresholds to historical graduation data to observe contrasts in historical graduation rates both above and below them. The historical effects of being over or under these thresholds on six-year graduation rates for FTF are represented visually in Figure 2 by URM / non-URM and Pell / non-Pell statuses. For URM students in the Fall 2012 – Fall 2014 cohorts, the six-year graduation rate increases from 20.8% to 68.1% above a First-term GPA of 1.84, and further increases to 78.2% above a First-term GPA of 2.89. For Pell-receiving students in the same cohorts, the six-year graduation rate increases from 21.3% to 69% above a First-term GPA of 1.82, and further increases to 79.7% above a First-term GPA of 2.94.

It is worth noting that both of the secondary First-term GPA thresholds for Pell and URM students shown above were fairly close to the *initial* First-term GPA thresholds that were found to significantly increase chances of graduation in the supplemental *four-year graduation* model for all FTF in the Fall 2012 – Fall 2016 cohorts: > 2.67 First-term GPA. This empirically illustrates that increases in First-term GPA that approach the secondary GPA threshold for increased odds of six-year graduation are *also more*

likely to increase the odds of graduation within four years. For this reason, we also display historical data in Figure 2 on changes in the four-year graduation rate for FTF both above and below the secondary First-term GPA thresholds noted above. As can be seen, the four-year graduation rate for URM students increases from 12.5% to 36.3% above a First-term GPA of 2.89; the four-year graduation rate for Pell-receiving students increases from 16.2% to 40.1% above a First-term GPA of 2.94.

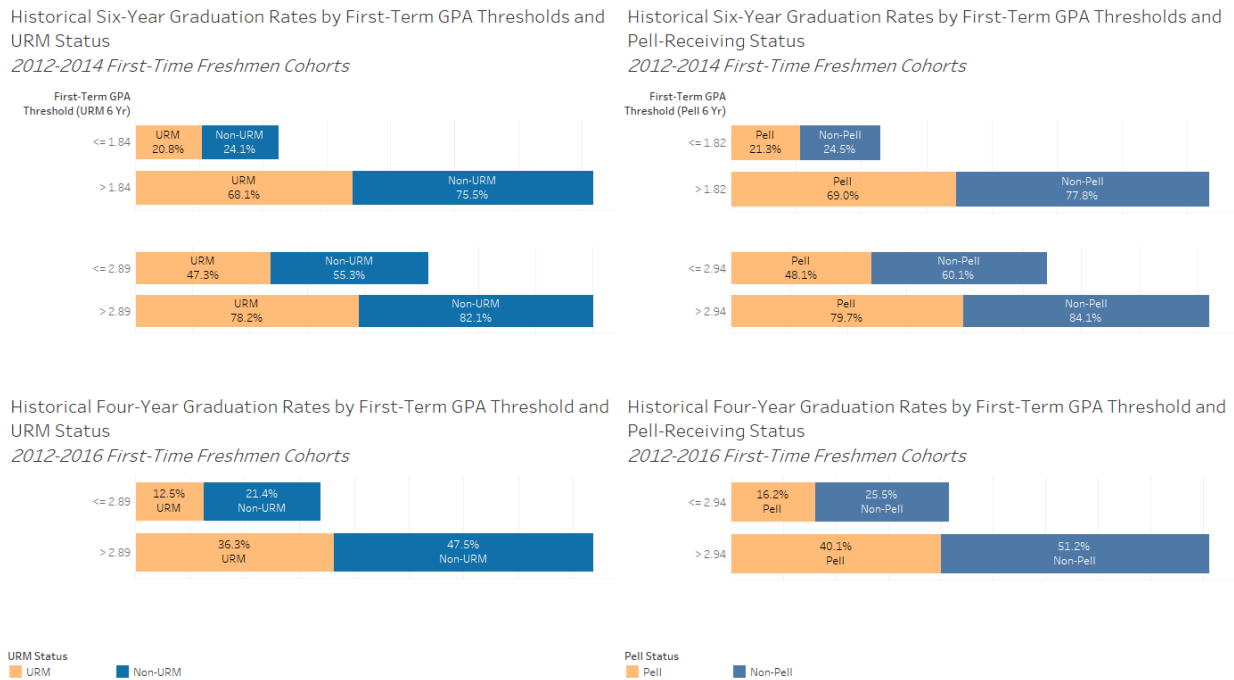


Figure 2. Historical Six- and Four-Year Graduation Rates by First-term GPA Thresholds for URM and Pell-Receiving Students

Entry GPA and College of Major

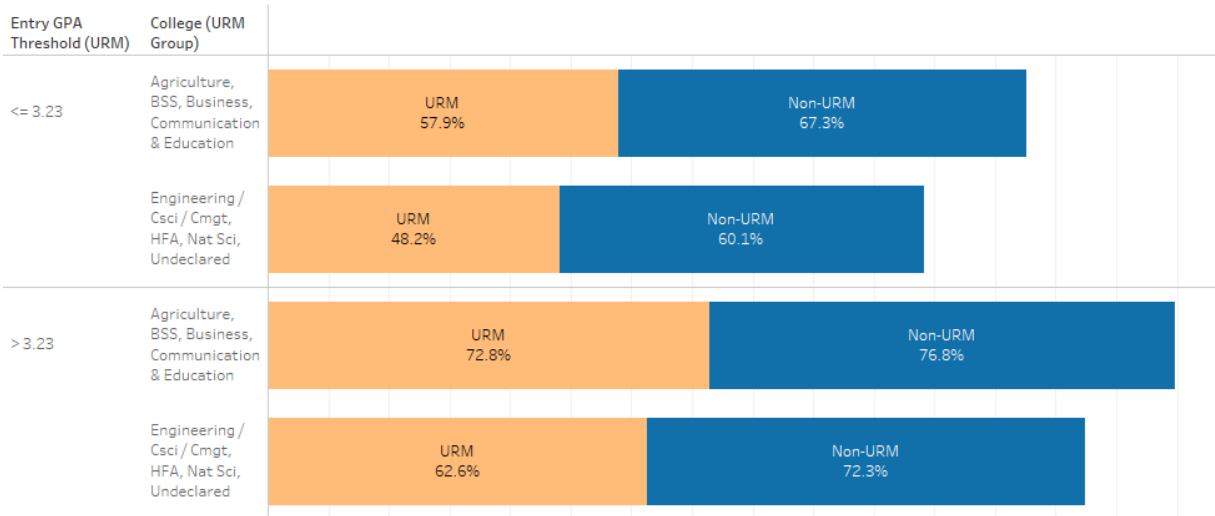
Besides First-term GPA, Entry GPA and College of Major were the most important secondary factors identified in the CRT and CHAID models for predicting graduation within six years. The decision tree models identified URM and Pell students as having an increased predicted probability of graduating within six years when their Entry GPA (in the case of first-time freshmen, high school GPA) was above a certain threshold, and when they had a declared major within particular colleges and not others. These factors are both notable because Entry GPA and College of Major can serve as *important early indicators* of students who may benefit from increased provision of support services: both are generally known after a student matriculates but before the start of their first term. The application of these identified Entry GPA thresholds and College groupings to historical six-year graduation rates for FTF by URM / non-URM and Pell / non-Pell statuses are shown in Figure 3.

- URM FTF
 - Entry GPA threshold: > 3.23
 - College of Major: Agriculture, Behavioral & Social Sciences, Business, or Communication & Education
- Pell-Receiving FTF

California State University Chico

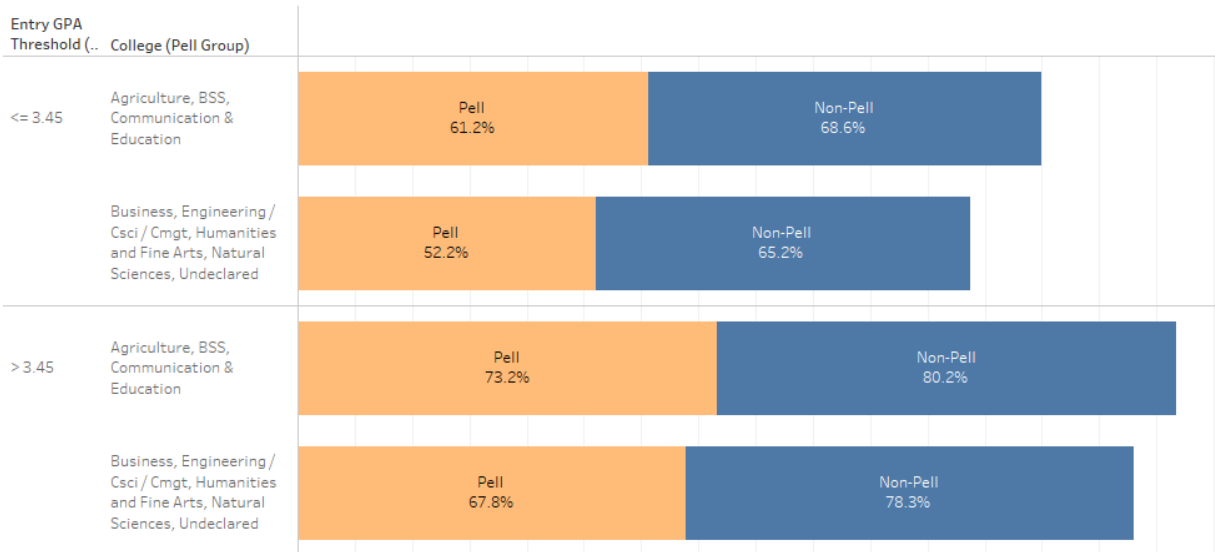
- Entry GPA threshold: > 3.45
- College of Major: Agriculture, Behavioral & Social Sciences, or Communication & Education

Historical Six Year Graduation Rates by Entry GPA Threshold, College Group, and URM Status
2012-2014 First-Time Freshmen Cohorts



URM Status
■ URM ■ Non-URM

Historical Six Year Graduation Rates by Entry GPA Threshold, College Group, and Pell-Receiving Status
2012-2014 First-Time Freshmen Cohorts



Pell Status
■ Pell ■ Non-Pell

Figure 3. Historical Six-Year Graduation Rates by Entry GPA and College of Major Groupings for URM and Pell-Receiving Students

California State University Chico

As can be seen, six-year graduation rates for URM FTF with declared majors in Agriculture, Behavioral & Social Sciences, Business, and Communication & Education increase from 57.9% to 72.8% above a 3.23 Entry GPA, while six-year rates for URM FTF with declared majors in Engineering / Computer Science / Construction Management, Humanities & Fine Arts, Natural Sciences, or who were undeclared increased from 48.2% to 62.6% above the same 3.23 Entry GPA threshold. Similarly, six-year graduation rates for Pell-receiving FTF with declared majors in Agriculture, Behavioral & Social Sciences, and Communication & Education increased from 61.2% to 73.2% above a 3.45 Entry GPA, while six-year rates for Pell receiving FTF with declared majors in Business, Engineering / Computer Science / Construction Management, Humanities & Fine Arts, Natural Sciences, or who were undeclared increased from 52.2% to 67.8% once above the same 3.45 Entry GPA threshold.

Note on Equity Gaps in Graduation Rates in Figures 2 and 3

In both Figures 2 and 3, readers should note that the effect of having a First-term GPA or Entry GPA that is above the respective threshold has a relatively similar magnitude of effect for URM / non-URM and Pell / non-Pell students – all become more likely to graduate with a higher First-term GPA, and their respective groups' graduation rates often increase by similar amounts once above the respective GPA threshold.

However, readers can also readily note that equity gaps in historical graduation rates *between* these student groups are relatively consistent, *both above and below* the First-term / Entry GPA thresholds and *within* the college groups. In other words, while increasing First-term GPAs benefit all students, and are of particular value to URM and Pell-receiving students in making graduation within six years more likely, they do not necessarily make academic outcomes more equitable *between* groups. This further underscores the importance of collecting systematic behavioral, social, and affiliation- / engagement-based data on students at the institutional level, as such data will likely aid Institutional Research and its partners in identifying factors that explain these inequitable outcomes between social groups.

Other Model-Identified Characteristics by Entry GPA Thresholds

Tertiary inputs identified by the CRT and CHAID models as important predictors of six-year graduation include *first-generation status and gender for Pell students*, and *gender and First-term attempted units for URM students*. Figures 4 through 7 show the breakdown of historical six-year graduation rates within Entry GPA thresholds for Pell and URM students for these student characteristics. It should be noted that gender-based disparities within historical six-year graduation rates that favor female students are significant, and cut across URM, Pell, and non-URM / non-Pell students.

As with the previous breakdowns of historical data shown above, Entry GPA can be used alongside these demographic and academic characteristics to coordinate provision of services and other interventions that may support enhanced student success in the first term. For URM students, data shown in Figure 7 also suggest the importance of attempting more than 12 units in the first term, but this must be balanced against the greater importance of keeping First-term GPA as high as possible to support graduation within six years.

Historical Six Year Graduation Rates by Entry GPA Threshold, First-Generation Status, and Pell-Receiving Status
2012-2014 First-Time Freshmen Cohorts

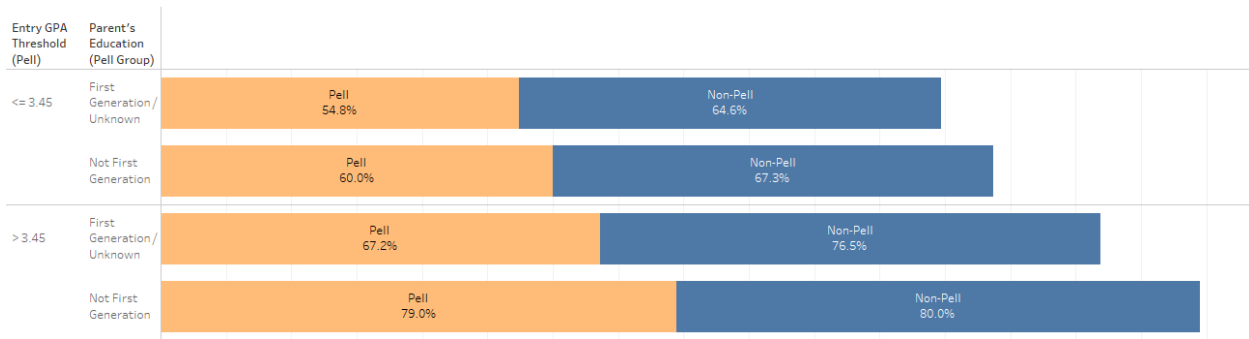


Figure 4. Historical Six-Year Graduation Rates by Entry GPA and First-Generation Status for Pell-Receiving Students

Historical Six Year Graduation Rates by Entry GPA Threshold, Gender, and Pell-Receiving Status
2012-2014 First-Time Freshmen Cohorts

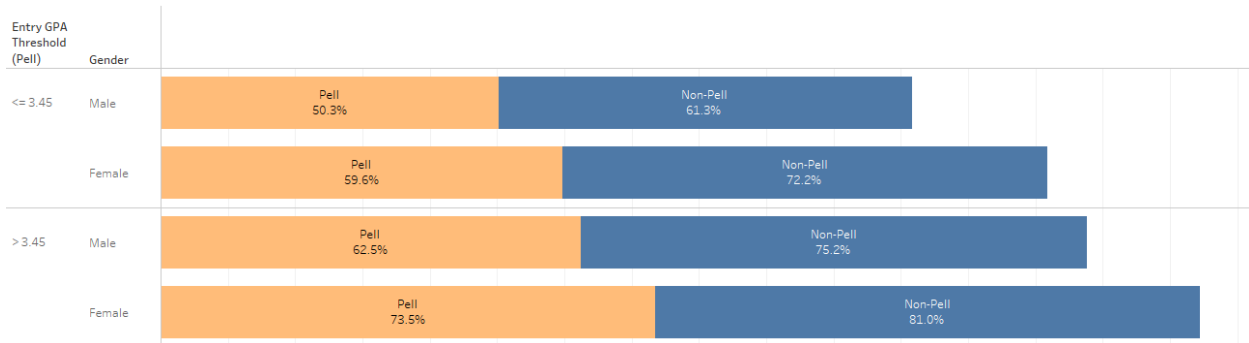


Figure 5. Historical Six-Year Graduation Rates by Entry GPA and Gender for Pell-Receiving Students

Historical Six Year Graduation Rates by Entry GPA Threshold, Gender, and URM Status
2012-2014 First-Time Freshmen Cohorts

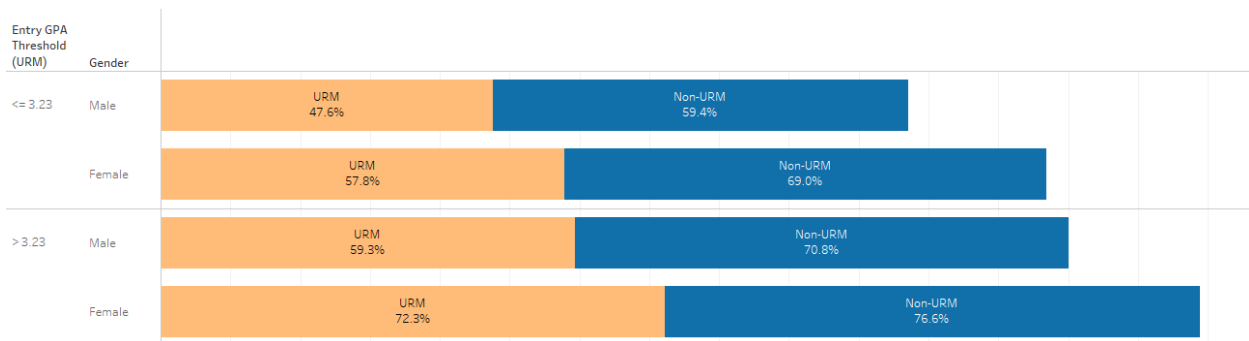


Figure 6. Historical Six-Year Graduation Rates by Entry GPA and Gender for URM Students

Historical Six Year Graduation Rates by Entry GPA Threshold, First-Term Attempted Unit Threshold, and URM Status
2012-2014 First-Time Freshmen Cohorts

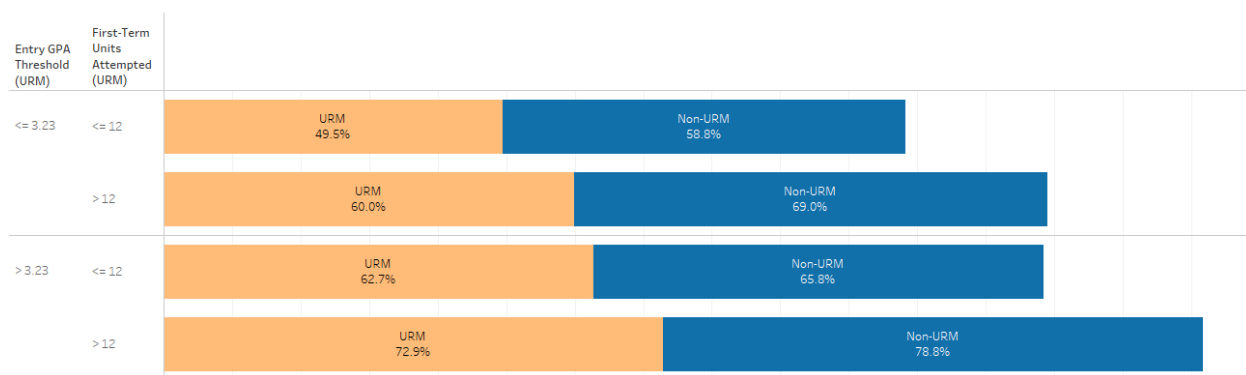


Figure 7. Historical Six-Year Graduation Rates by Entry GPA and First-term Units Attempted Threshold for URM Students

Further Discovery

Institutional Research staff will continue to refine and update these models and results as improved institutional data become available. Two factors that IR staff are currently introducing into these models include STEM Major status and whether or not a student attended orientation before the start of classes. Other factors of interest include segmenting Pell-receiving students by relative level of financial needs, further disaggregating first-generation students by level of parental education, and introducing data on DFW rates as an input. IR staff are also working to segment models by the recency of cohort terms and by different academic outcomes. E.G., to see if different sets of input factors are found to be useful predictors of one- and two-year retention outcomes, versus those identified as predictors of four- and six-year graduation for earlier cohorts. As models are further refined, and as the campus makes changes to underlying social and academic conditions, some variability in rankings of input importance will become likely. For example, First-term GPA and Entry GPA may not persist as the most significant predictors of six-year graduation, and other factors may rise to the top.

Once more, the importance of behavioral, social, organizational affiliation, and engagement-based data on students cannot be emphasized enough for successfully modeling academic outcomes such as retention and graduation, as the omission of these factors due to data availability is perhaps the single-greatest limitation of the models presented in this report. For instance, one could make a strong hypothesis that major influencers of First-term GPA (and by extension, six-year graduation rates) might include the average number of hours per week that a student works at an off-campus job, the quality of their experiences with advising or tutoring services, their organizational affiliations, or their relative level of engagement with support services. It is critically important for the Chico State community to consider means by which to collect these types of data more thoroughly at a systematic level, both through existing instruments such as the National Survey of Student Engagement (NSSE) and Student Evaluation of Teaching (SET) forms, as well as by creating new, empirically-supported means for their collection at the college and departmental levels.