

# Student Success Predictive Model Report – uwm\_10132

---

August 26, 2020

## Executive summary

EAB has built a customized Student Success Predictive Model (SSPM) for your institution that predicts the persistence likelihood of your students. Your SSPM incorporates the latest breakthroughs in statistics and data science, placing your institution at the cutting edge of student-insight technology. It is a powerful tool for promoting your students because it gives you invaluable insight into their likelihood of academic success. This document provides an overview of the SSPM, describes how it was built and extensively customized and optimized for you, and details benchmarks of its predictive performance.

### Performance Summary

The primary metric EAB uses to benchmark model performance is high-risk student identification rate. It is based on the most common use case for the model: that you are designing a campaign targeting high-risk students but only have the capacity to advise a limited subset of your total student population. In this case, your goal is to efficiently use your constrained resources to reach as many of your school's actual high-risk students as possible.

The table below summarizes your SSPM's performance and compares it to the following notional models:

- A fictitious, perfectly prescient model (Crystal Ball).
- A model based exclusively on students' cumulative GPAs (GPA Model).
- A model that randomly targets students (Blind Campaign).

The columns assume different percentages of your total student population that you are able to cover in the campaign, while the rows provide the percentage of your school's actual high-risk students that will be identified in the campaign based on each model.

The bottom row highlights the substantial relative percentage gains achieved in going from the simple GPA Model to your advanced Student Success Predictive Model and demonstrates that your model is much better at distinguishing between students who are on track to graduate and those that need intervention in order to succeed.

Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	20%	40%	98%	100%
Your Model	14%	25%	47%	75%
GPA Model	9%	16%	35%	63%
Blind campaign	5%	10%	25%	50%
<b>Relative Percentage Gain</b>	<b>56%</b>	<b>56%</b>	<b>34%</b>	<b>19%</b>

Your SSPM is high-performing; it can be used confidently to both assess individual students and efficiently design effective, targeted intervention campaigns.

## Introduction

### Overview

This document provides information about your institution’s custom Student Success Predictive Model (SSPM). It describes how the model was built; details the top success indicators or “predictors” used in the model and provides metrics characterizing the predictive power of the model.

The SSPM uses your school’s student records to predict the likelihood that any chosen student will persist to the next term of the regularly scheduled academic year (or graduate before then). This is done by first “training” a statistical model using the records of historical students in order to determine—and assign values to—the items derived from those records that are “predictors” of persistence outcomes.

The model outputs a success score between zero and one estimating the probability that a selected student will persist to the next term. That is, each student’s success score corresponds to the model’s estimate of their likelihood of persisting to the next term. Since it is not possible to build a perfectly prescient model, it is important to state that a score of one does not guarantee a student’s persistence. Nor does a score of zero guarantee their failure. A success score of 0.7 for instance, may be interpreted as our expectation that, on average, seven of ten students with this score will persist to the next term.

### Methodology

The SSPM uses the latest advances in data science to estimate persistence likelihood for each student, from incoming freshmen to nearly-graduating seniors. A customized set of predictors are constructed from student records, and then combined and weighted using an automated training process. EAB’s Data Science team customizes this process for each partner, and uses a variety of optimization tools to ensure the best possible performance given the data available.

As described above, the model is trained from recent historical student records; in particular, students satisfying the following criteria were used:

- Matriculated between 2006-08-26 and 2018-09-04.
- Had at least one registered term.
- Were seeking a degree.

Technical details: The model is a combination of several penalized logistic regression models applied to different subgroups of students. The predictors include simple lookups of student records (e.g., high school GPA), as well as composite attributes derived from them whose details are proprietary.

## Your Institution's Model

The SSPM includes a wide variety of success indicators called "predictors" in order to ensure maximal predictive power. We use your institution's historical data to determine the best set of predictors that most accurately reflects the underlying patterns of your students. The items below were found to be good predictors for your institution. The predictors in these lists are not equally important and may not be the same for all subgroups; the statistical model learns how to identify and assign values to the best predictors for each subgroup of students. For instance, we might expect high school GPA to be highly relevant for freshmen, but minimally important for seniors.

## Your Predictors

The lists below describes the predictors for each subgroup of students in the model. We are sharing these predictors to help you understand how the model works and the types of variables that are predictive of success at different points in a student's academic career. Knowing these variables can help build understanding of the model and may provide insight into where to start conversations with different groups of students. However, multivariate machine learning models are very complicated, and sometimes unintuitive, so the individual variables should be interpreted cautiously. The SSPM is designed to maximize predictive accuracy, not to maximize our understand the impact of any individual input variable. This is not the same as a controlled study on the influence of these variables, and the inclusion of any variable on this list does not imply a causal effect. Many variables are highly correlated with one another, and therefore "High Impact Predictors" may change from one model iteration to another even if the training data and model structure are similar. Therefore, EAB does not recommend using this list to drive specific actions around individual variables except when used in conjunction with external studies on causality and, of course, the human intelligence of subject matter experts.

For each sub-model, the predictors are organized into two sections: "High Impact Predictors" and "Other Predictors." "High Impact Predictors" are the predictors that are responsible for more than 5% of the variance in scores across all of the students in a credit bin. This may mean that the variable has a moderate impact on the scores of many students, or a high impact on the scores of just a few students. Some variables have a significant impact on the students they affect, but affect only a low number of students and therefore do not count as "High Impact Predictors." Just because a variable is or is not a "High Impact Predictors" for the population does not mean that it is or is not an important factor for an individual student. The "Other Predictors" are variables that the model identified as statistically significant predictors, but are responsible for less than 5% of the variance in scores across the population. They may be mildly predictive for many students or highly predictive for a very low number of students.

Although the "Other Predictors" are individually weak predictors, collectively they are responsible for a significant portion of model performance.

In the following section, we enumerate the "High Impact Predictors" for each sub-model. To see the full list of "High Impact" and "Other" Predictors for each sub-model, please refer to Appendix III near the end of the report.

- **Pre-Enrollment Students High Impact Predictors**

Admit Code

High School Percentile

Race/Ethnicity

Transfer Indicator

High School GPA

...and 5 low impact predictors.

- **Day 1 Students High Impact Predictors**

Admit Code

Credits Attempted Current Term

Race/Ethnicity

High School Percentile

Average Credits Attempted per Term

Average Success Outcome of Students Declared in Same Major

High School GPA

...and 17 low impact predictors.

- **Students with Between 1-60 Accumulated Credits High Impact Predictors**

Credits Attempted Current Term

Number of Completed Terms

Cumulative GPA

First Term Transfer Credits

Average Success Outcome of Students Declared in Same Major

...and 22 low impact predictors.

- **Students with Between 61-120 Accumulated Credits High Impact Predictors**

Credits Attempted Current Term

A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.

Average Credits Attempted per Term

Admit Code

Cumulative GPA

Ratio of Earned to Attempted Credits

...and 21 low impact predictors.

- **Students with More Than 120 Accumulated Credits High Impact Predictors**

Cumulative GPA

Credits Attempted Current Term

Average Success Outcome of Students Declared in Same Major

Average Credits Attempted per Term

Proportion of Transfer Credits

Ratio of Earned to Attempted Credits

Age at First Term

...and 19 low impact predictors.

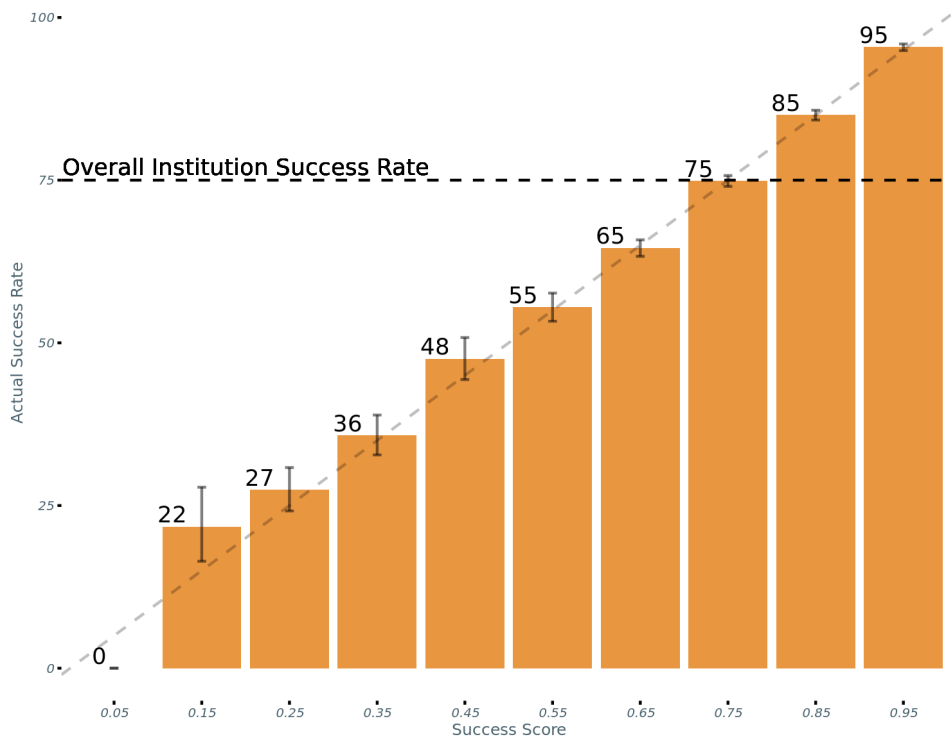
## Model Performance

Your SSPM is well-calibrated and its performance has been thoroughly characterized using a “test set” of your historical students that was set aside from the training set NA Blind Campaign model that randomly targets students. This section describes the most insightful performance benchmarks and compares your SSPM to these other notional models.

### Calibration

Calibration offers an intuitive way to evaluate a model by capturing how close its estimated probability scores are to reality.

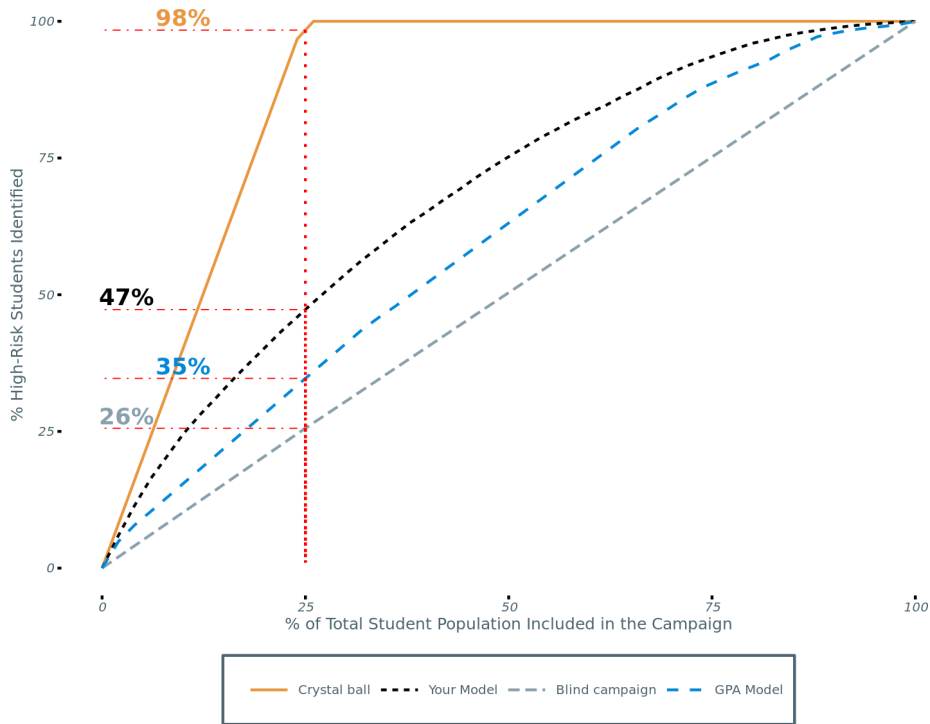
Students are divided into different bins along the horizontal axis according to their success score, while the vertical height of each bin indicates the actual persistence rate of the historical students it contains. The horizontal line shows the overall percentage of students that persisted to the next term.



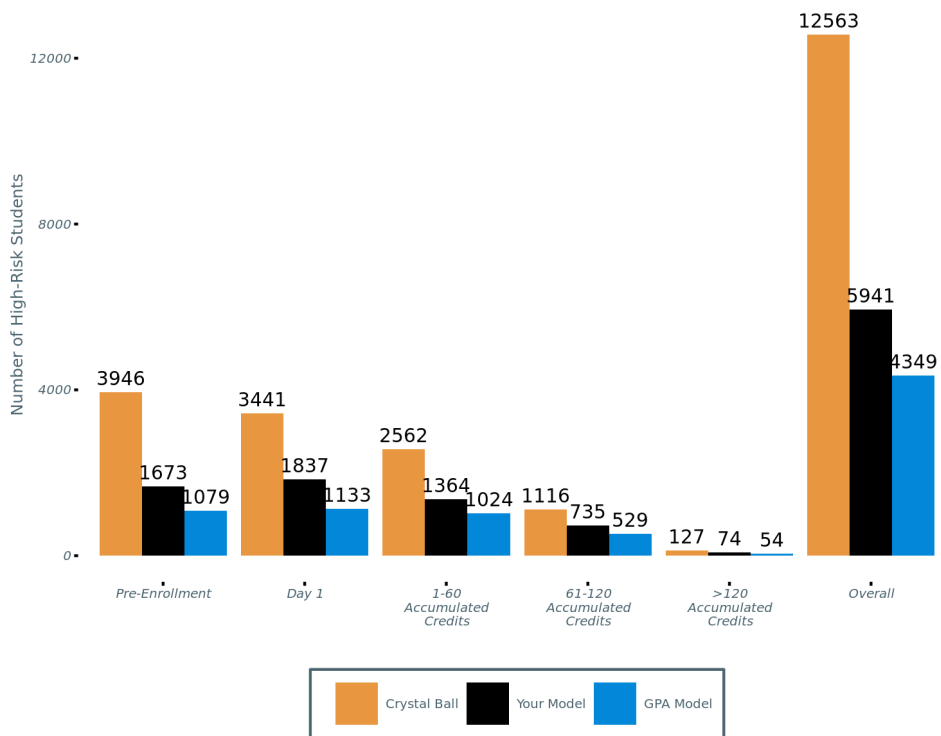
## High-Risk Student Identification Rate

The SSPM enables you to rank students by order of risk (i.e., success scores from low to high) so that you can most efficiently target as many high-risk students for intervention as your institution or office/department can effectively handle. Let’s assume, for instance, that you are designing an intervention campaign targeting high-risk students and have the capacity to advise N students. Let’s assume you use different predictive models to generate lists of N targeted high-risk students, and step forward in time to compare their performance by evaluating the percentage of those N students that did not persist to the next term. This performance comparison is summarized in the high-risk student identification rate chart below, which shows the percentage of actual high-risk historical students (i.e., students that did not persist to the next term.) identified by the model vs. the percentage of the total student population targeted in the campaign. For example, if you design a campaign that includes 25% of your total student population, then the percentage of your school’s high-risk students identified by the campaign will be 98, 47, 35, and 26 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

High-risk student identification rate provides a powerful and transparent performance benchmark of model performance; the large performance enhancement gained in going from the simple GPA Model to your advanced SSPM is clearly visible in the chart.



High-risk student identification rates can also be converted to actual numbers of students and compared across different accumulated credit subgroups, as shown in the figure below for campaigns targeting 25% of the total student population.

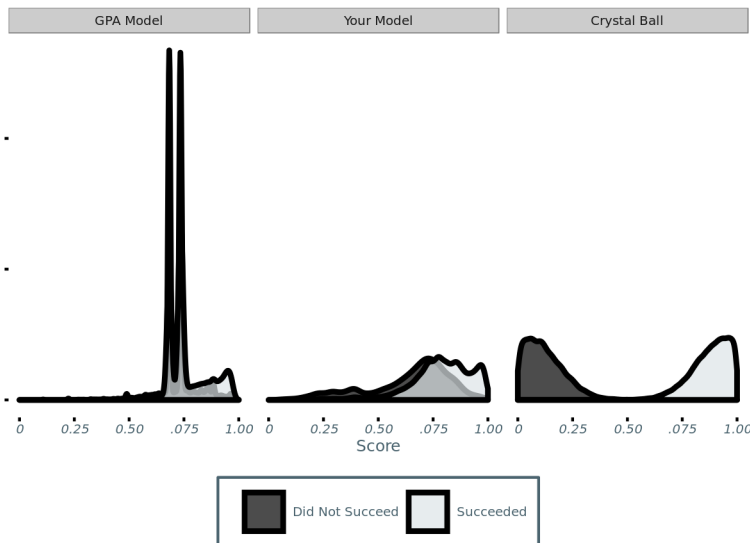


## Lift

We may divide the percentage of actual high-risk students identified by a given model by the percentage found by a Blind Campaign to create a new metric called “lift”. For instance, a lift of two would mean that a campaign based on your SSPM identified twice as many high-risk students as a Blind Campaign, while a lift value less than one would indicate that your model identified fewer actual high-risk students than simply choosing from your student population at random. Considering a campaign that includes 25% of your total student population, lift is 4.01, 1.89, 1.5, and 1.00 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

## Separation

Displaying the distributions of success scores for students in the historical test set who did and did not persist to the next term also provides an intuitive sense of a model’s performance. We see in the charts below that successful students (light gray) typically have higher scores than unsuccessful ones (dark gray) for both your SSPM and the GPA Model, as you would expect, but that your SSPM is much better at separating these two student populations from each other. That is, the graphic demonstrates that your SSPM ascribes high success scores to successful students and low success scores to unsuccessful students more accurately than the GPA Model. A perfect prediction would result in complete separation between the students (shown in the Crystal Ball chart on the right).



## Conclusion

The performance of your institution’s Student Success Predictive Model has been extensively optimized and evaluated; the model will provide your school and its advisors with invaluable and otherwise unobtainable insight into your students’ likelihood of academic success. The model incorporates the latest breakthroughs in statistics and data science and places your institution at the cutting edge of student-insight technology. Your advisors may use it with



confidence to both assess individual students and design effective and efficient targeted campaigns.

## Appendix I: Evaluating AUC

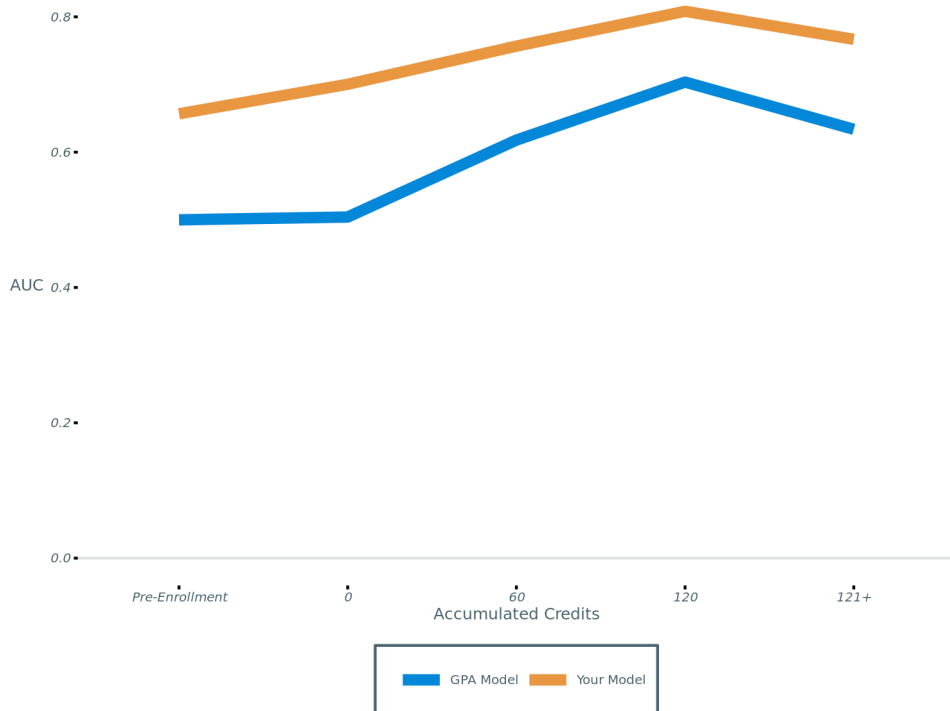
We commonly use AUC to measure and tune the performance of your Student Success Predictive Model across your institution's entire student population and different subgroups. AUC stands for Area Under the Curve and is a measure used extensively in data science, which ranges from 0.5 (pure chance) to 1.0 (Crystal Ball). We evaluate your SSPM's AUC in comparison to the notional GPA Model; your SSPM's larger AUC indicates that it identifies high-risk students more accurately than the GPA Model. This is the type of rule-of-thumb based approach that academic advisors intuitively know is useful.

The table below shows AUC values for your SSPM and the GPA Model.

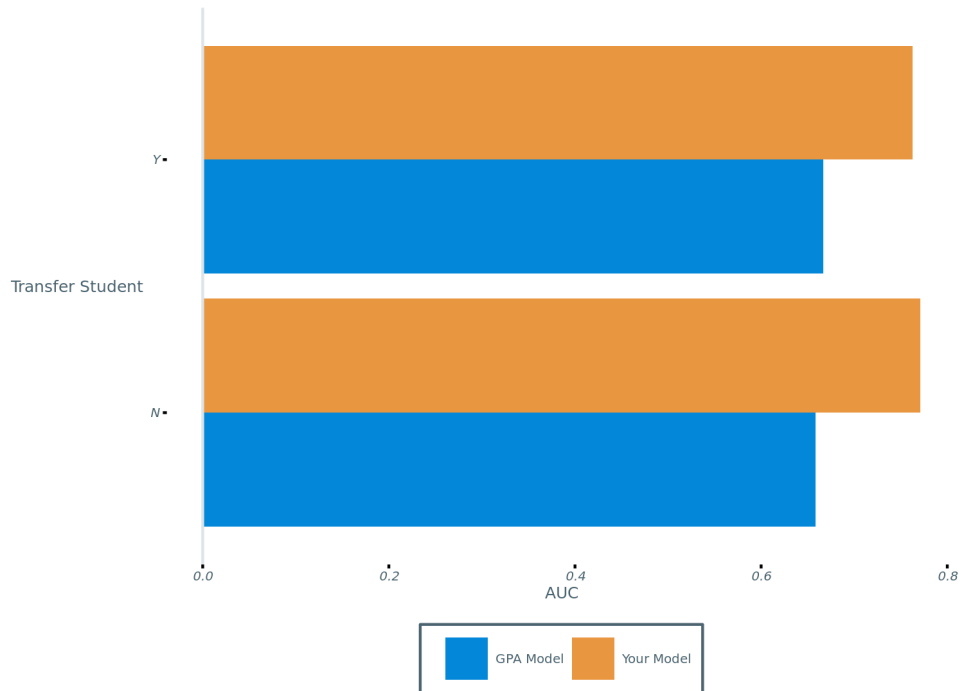
Model	AUC
GPA Model	0.67
Your Model	0.77

As part of validating your SSPM, we examine subgroups of students to ensure that it consistently performs. The figures below show the AUC values for students with different levels of accumulated credits and for Transfer/Non-Transfer students.

### AUC for Students with Different Numbers of Accumulated Credits



## Accuracy for Transfer/Non-Transfer Students

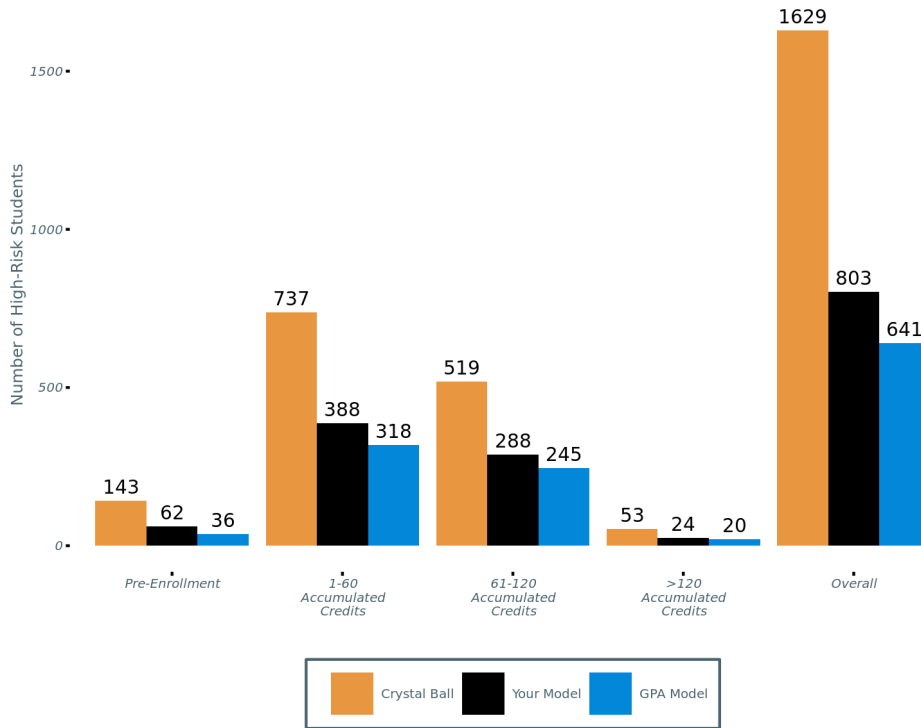


## Appendix II: High-Risk Student Identification Rate for Murky Middle and Top Performing Students

Your Student Success Predictive Model's performance varies across different subgroups of students. This appendix provides plots and tables evaluating model performance in terms of high-risk student identification rate for two student subgroups: Murky Middle and Top Performing. The same plots provided for the overall student population in the main body are shown in this appendix for two student subgroups.

### Murky Middle

Murky Middle students are defined as those students whose cumulative GPAs are between 2.0 and 3.0.

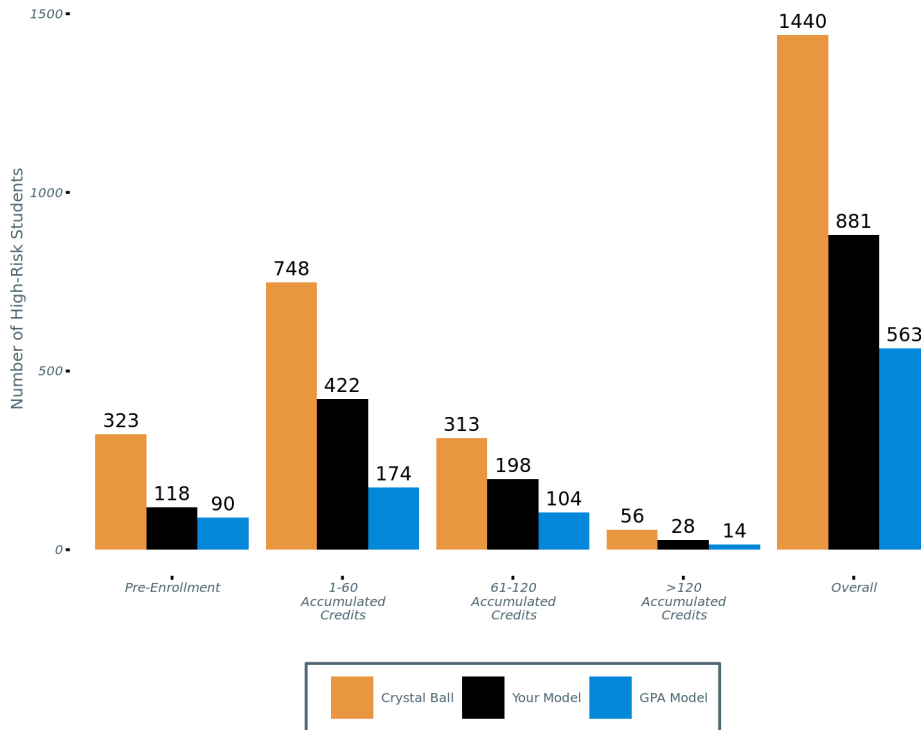


Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	22%	44%	100%	100%
Your Model	13%	25%	49%	77%
GPA Model	11%	19%	39%	68%
Blind campaign	5%	10%	25%	50%
<b>Relative Percentage Gain</b>	<b>18%</b>	<b>32%</b>	<b>26%</b>	<b>13%</b>

### Top performing students

Top performing students are defined as those students whose cumulative GPAs are greater than 3.



Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	40%	81%	100%	100%
Your Model	19%	34%	61%	86%
GPA Model	9%	17%	39%	77%
Blind campaign	5%	10%	25%	50%
<b>Relative Percentage Gain</b>	<b>111%</b>	<b>100%</b>	<b>56%</b>	<b>12%</b>

## Appendix III – Additional Predictor Information

### All Predictors

The list below enumerates all predictors used in each submodel, including “Other Predictors” that were not important enough to be included in the “Your Predictors” section of the report.

- Pre-Enrollment Students High Impact Predictors
  - Admit Code
  - High School Percentile
  - Race/Ethnicity
  - Transfer Indicator

- High School GPA
- **Pre-Enrollment Students Other Predictors**
  - International Indicator
  - First Generation Indicator
  - Gender
  - In State Resident Indicator
  - High School Size
- **Day 1 Students High Impact Predictors**
  - Admit Code
  - Credits Attempted Current Term
  - Race/Ethnicity
  - High School Percentile
  - Average Credits Attempted per Term
  - Average Success Outcome of Students Declared in Same Major
  - High School GPA
- **Day 1 Students Other Predictors**
  - First Generation Indicator
  - International Indicator
  - Ratio of Credits Attempted Current Term to Prior Term
  - SAT/ACT Verbal Score Percentile
  - Gender
  - SAT/ACT Math Score Percentile
  - High School Size
  - In State Resident Indicator
  - Age at First Term
  - Cumulative GPA
  - Ratio of Earned to Attempted Credits
  - A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.
  - Number of Completed Terms
  - Proportion of Transfer Credits
  - Trend in Term GPA

- Estimated Skills
- **Students with Between 1-60 Accumulated Credits High Impact Predictors**
  - Credits Attempted Current Term
  - Number of Completed Terms
  - Cumulative GPA
  - First Term Transfer Credits
  - Average Success Outcome of Students Declared in Same Major
- **Students with Between 1-60 Accumulated Credits Other Predictors**
  - Average Credits Attempted per Term
  - Proportion of Transfer Credits
  - A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.
  - Admit Code
  - High School Percentile
  - Estimated Skills
  - Trend in Term GPA
  - Race/Ethnicity
  - International Indicator
  - Transfer Indicator
  - Gender
  - Ratio of Earned to Attempted Credits
  - Recent Change in GPA
  - In State Resident Indicator
  - First Generation Indicator
  - SAT/ACT Verbal Score Percentile
  - High School GPA
  - High School Size
  - Age at First Term
  - SAT/ACT Math Score Percentile
  - Ratio of Credits Attempted Current Term to Prior Term
- **Students with Between 61-120 Accumulated Credits High Impact Predictors**
  - Credits Attempted Current Term

A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.

Average Credits Attempted per Term

Admit Code

Cumulative GPA

Ratio of Earned to Attempted Credits

- **Students with Between 61-120 Accumulated Credits Other Predictors**

Average Success Outcome of Students Declared in Same Major

Trend in Term GPA

Number of Completed Terms

Age at First Term

Recent Change in GPA

High School Percentile

SAT/ACT Verbal Score Percentile

Ratio of Credits Attempted Current Term to Prior Term

Gender

Proportion of Transfer Credits

First Generation Indicator

Race/Ethnicity

Estimated Skills

International Indicator

Transfer Indicator

First Term Transfer Credits

High School Size

SAT/ACT Math Score Percentile

High School GPA

In State Resident Indicator

- **Students with More Than 120 Accumulated Credits High Impact Predictors**

Cumulative GPA

Credits Attempted Current Term

Average Success Outcome of Students Declared in Same Major

Average Credits Attempted per Term

Proportion of Transfer Credits

Ratio of Earned to Attempted Credits

Age at First Term

- **Students with More Than 120 Accumulated Credits Other Predictors**

Trend in Term GPA

Race/Ethnicity

SAT/ACT Verbal Score Percentile

Admit Code

A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.

Gender

Ratio of Credits Attempted Current Term to Prior Term

High School Percentile

High School GPA

In State Resident Indicator

SAT/ACT Math Score Percentile

International Indicator

Estimated Skills

First Term Transfer Credits

Number of Completed Terms

Recent Change in GPA

First Generation Indicator

High School Size

Transfer Indicator

### **Predictor Descriptions**

The list below provides detailed descriptions of all the predictors used in your model. We discussed the most important among these in the "Your Predictors" section of the report. This list is ordered alphabetically.

- A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.: A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major. This percentile score ranks students in comparison to the performance of their peers' in the same major; e.g., a sociology student with a score of 80 has a higher cumulative GPA than 80% of all students declared in the sociology



major. Students declared in multiple majors are assigned a percentile value that corresponds to the mean average of their scores for each major.

- Admit Code: A student's admission type (i.e., first time freshman, first time transfer, conditional admit, etc.)
- Age at First Term: A student's age upon starting their first term at your institution.
- Average Credits Attempted per Term: The average number of credits a student has attempted per term.
- Average Success Outcome of Students Declared in Same Major: This score indicates the average success outcome of all students enrolled in a given student's chosen major. E.g., if the model's success outcome is whether a student eventually graduates, and 90% of chemistry students do, then the score will be 90% for all chemistry students. Students declared in multiple majors, however, are assigned the mean average score across all of their majors.
- Credits Attempted Current Term: The number of credits a student is attempting in the current regular term. (The number of credits a student attempted in the most recent regular term is used in the case that a regular term is not currently in session.)
- Cumulative GPA: A student's cumulative GPA.
- Estimated Skills: A student's estimated academic skills. More specifically, we identify underlying patterns in the grades students earn in different courses – e.g., some students may have a history of excelling in math-related courses but not writing-related courses – and call the discrete factors behind these patterns "skills".
- First Generation Indicator: "Yes" or "No" indicator of whether any of an individual's parents have ever earned a bachelor's degree.
- First Term Transfer Credits: The number of credits a student transferred from other institutions upon matriculation.
- Gender: A student's gender.
- High School GPA: A student's high school GPA.
- High School Percentile: A student's high school rank in terms of percentile.
- High School Size: The size of an individual's high school student body.
- In State Resident Indicator: A "Yes" or "No" indicator of whether a student is a resident of your institution's home state.
- International Indicator: "Yes" or "No" indicator of whether an individual is an international student.
- Number of Completed Terms: The number of terms a student has completed at your institution.
- Proportion of Transfer Credits: The proportion of a student's credits that were earned at another institution.
- Race/Ethnicity: A student's race/ethnicity.

- Ratio of Credits Attempted Current Term to Prior Term: The number of credits a student attempted in the current regular term as compared to the number of credits they attempted in the prior regular term. (The most recent regular term and the one prior to it are used in the ratio in the case that a regular term is not currently in session.)
- Ratio of Earned to Attempted Credits: The overall number of credits a student has earned divided by the number of credits they have attempted.
- Recent Change in GPA: The difference in a student's GPA from the prior two complete terms
- SAT/ACT Math Score Percentile: A student's highest percentile achieved in either the SAT or ACT math test. We calculate a student's math percentile as the highest percentile they earned in either the SAT or ACT math tests. A percentile score ranks students in comparison to their peers' performance; e.g., a percentile score of 80 indicates that a student outperformed 80% of his peers in either the SAT or ACT math tests.
- SAT/ACT Verbal Score Percentile: A student's highest percentile achieved in either the SAT or ACT verbal test. We calculate a student's verbal percentile as the highest percentile they earned in either the SAT or ACT verbal tests. A percentile score ranks students in comparison to their peers' performance; e.g., a percentile score of 80 indicates that a student outperformed 80% of his peers in either the SAT or ACT verbal tests.
- Transfer Indicator: "Yes" or "No" indicator of whether the student transferred from another institution.
- Trend in Term GPA: A measure of the change over time in a student's term GPAs.