

Understanding the Risk Score

A Guide to How EAB Builds their Models and Risk Scores

Predictive Models and Risk Score

A predictive model takes into consideration all the variables in a given data set in order to make informed predictions about what might happen in the future. At EAB, our predictive model takes 10 years of historical student data to suggest the likelihood of various outcomes, like four-year graduation, six-year graduation, any graduation, next-term persistence, or next fall persistence. It is not a guarantee of these outcomes. The purpose of using predictive modeling is to assist institutions and their students to make informed choices to meet their diverse needs and reach their educational goals with real data.

Across our member institutions, the Student Success Predictive Model performs at a high level of accuracy of predicting graduation outcomes based on the available information. We use academic, demographic, and other pre-enrollment data in order to capture as much information as possible about a student's academic ability, performance, and preparation to determine their likelihood of success. The risk model analyzes 40+ predictive features, shown from our research to be strong indicators of graduation success.

Why Use Predictive Models?

There are four major benefits EAB has found in using predictive models and risk scoring

- **Visibility:** Predictive models show where risk is distributed across your institution. They also surface hidden pools of risk among your student population
- **Early Intervention:** By using Day 1 risk factors as well as early risk predictions, institutions can intervene with students earlier in their educational paths
- **Triage:** Institutions can find and intervene with the highest-risk students in their cohorts in terms of risk. In addition, risk scores help identify which students to target first
- **Impact:** Using these models, institutions can see whether or not students are improving thanks to interventions and ideally, reduce pools of risk in the student population

What Are Influencers and How Do They Affect Risk Scores?

An *influencer* is an actual or close approximation of a variable that gets fed into the model from your data. Variables undergo a machine learning process. Machine learning is the science of getting computers to learn and act in a human fashion, where computers improve their learning over time by being given more data and information over time. As your institution's data is processed through machine learning, variables are assigned a weight. The "weight" of a variable is a way to designate its relative importance on student outcomes in your data and will depend on your institution's individual situation and student population. For example, if your institution has a large population of transfer students, the model will weight "Pre-Enrollment Data" more heavily. If your school has fewer transfer students, the model may weight high school GPA more heavily.

Influencers are a set of variables that represent a part of the student risk score that we think can be attributed to performance in an area. They represent a subset of predictors – one of 60 variables that our models find correlation between that variable and student success. The point of these influencers is to distill down an array of variables to the few things that are truly having an impact on students.

For example, grades often predict student success. However, student grades can then be broken down into many variables like cumulative GPA, mix of grades, the increase or decrease in grades in the current term compared to the previous term, and so forth. The model can further transform student grades to find relevant differences – two identical GPAs could discover that a student whose 3.0 GPA reflects all Bs is more likely to be successful compared to a student with an even mix of As and Cs.

Our data science team provides some “human curation” for the list of influencers displayed on the student’s profile page by selecting the variables most likely to be actionable by advisors and excluding variables that are solely mathematical.

EAB then sets thresholds for each member institution to determine if the impact correlates to positive, neutral, or negative impact. A neutral impact means the factor is present in the model, but not of significant impact either way. We include neutral impact scores to be transparent and show more influencers, even if they have no impact. The scores and thresholds used are always in comparison to the “average” student as created from historical student data.

List of Influencers

This is a list of influencers that EAB displays on the student profile. Not every feature applies to every member and may not be included in your institution’s risk score if so. In addition, not every factor used within your institution’s predictive model will display on the student profile page.

- Cumulative GPA
- GPA Trend
- High School GPA
- Transfer Student
- Average Outcome in Major
- Percentile Rank in Major
- Earned to Attempted Credit Ratio
- Lifetime Accumulated Credits

How Do I Use My Risk Score/Predictive Model with Students?

Influencers can serve two major purposes. One can lead advisors and faculty to find actionable opportunities for students. The other is to explain the reasons for a student risk, like pre-enrollment student data. While “explainer” data may not be actionable, it can help students and advisors to understand their risk score OR help advisors discover risk influencers in advance for earlier intervention.