Sam Houston
State University

Data Analytics and Decision Support

Flipping the Script on Early Alerts

Tristan Young
Melissa Villarreal
Duleepa Sembukutti
Brandon Cooper

OUTLINE

1 Background

- Early Alert system gaps
- Modeling for term outcomes
- Policy changes & emerging demands

2 The Pivot

- Heterogeneity in the LMS gradebook
- Developing an Engagement Adjusted Mean
- Identifying at-risk students

3 The Tool

- Co-designed with faculty & the center for faculty development
- Insights into student support level across entire course schedule

4 Future Directions

- Predictive capacity of EAM
- Modeling with other predictors

BACKGROUND

Early Alert Gaps at Sam Houston State University

- **48.7% of first-time freshmen** who ended Fall in unsatisfactory standing were **not identified by early alert system**
- Initial goal: Identify at-risk students **earlier** and more **comprehensively**

Core Question:
Can predictive analytics strengthen early intervention?

CROSS-FUNCTIONAL COLLABORATIONS

Data Analytics & Decision Support

Student Affairs

Academic Affairs

Academic Success Center (ASC)

Teaching & Learning Center

The Models & the Tool

We developed **separate models for**:

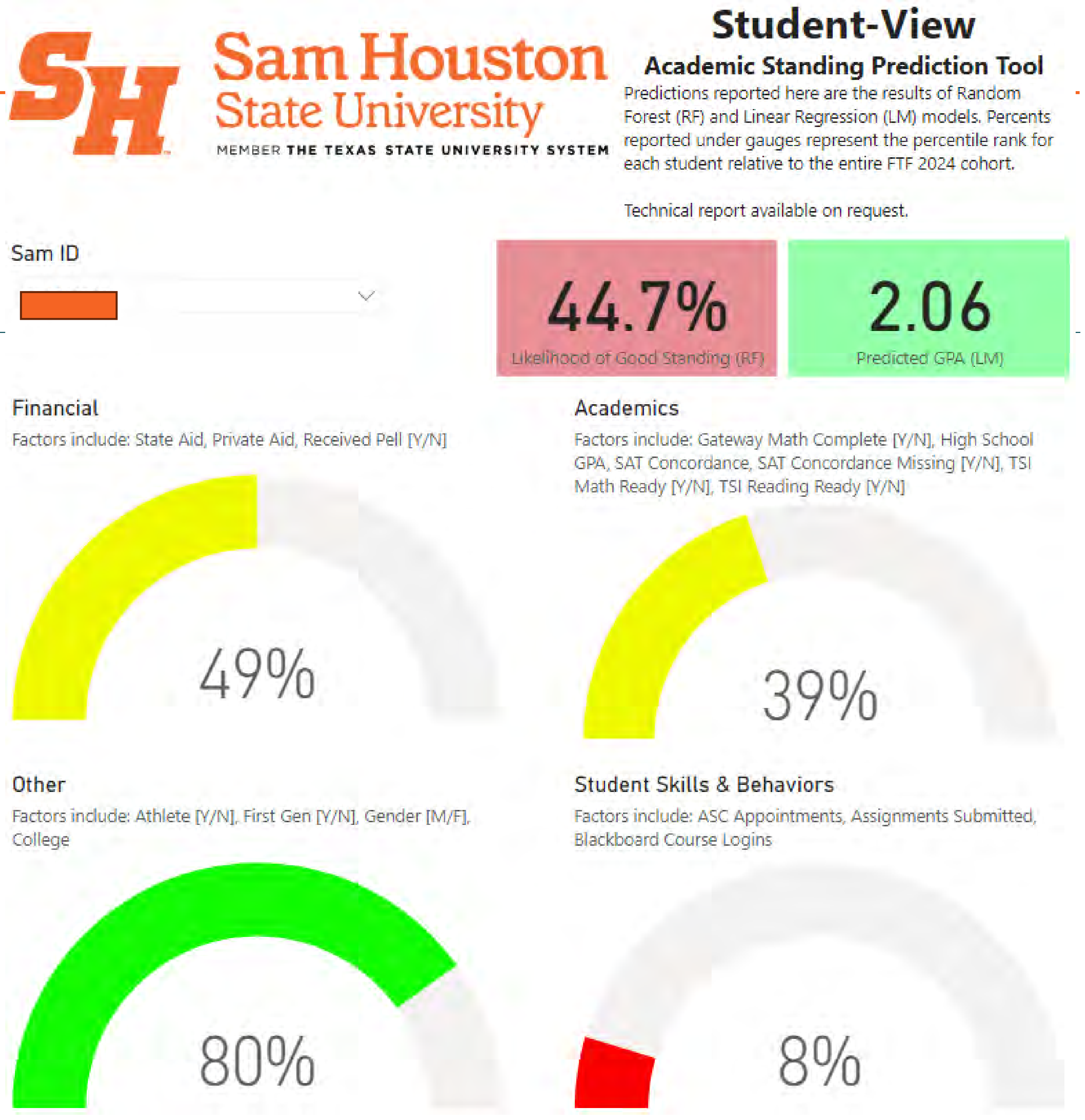
- First-Time Freshmen (FTF)
- First-Time Transfers (TR) - Beginning Fall 2025
- Continuing Freshmen (CF) - Beginning Fall 2025

Each group got **two models:**

1. Random Forest (Classification)

2. Linear Regression (GPA Prediction)

Together: **Who is at risk + Why**



ACADEMIC STANDING PREDICTIVE TOOL

LIST VIEW

Sam Houston State University
MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

List-View

Academic Standing Prediction Tool

Predictions reported here are the results of Random Forest (RF) and Linear Regression (LM) models. Table results can be exported by hovering over the table and clicking on the ellipses representing "More Options."

Technical report available on request.

Sam ID
All

Good Standing (Hold CTRL to Multiselect)
All

Alerted?
All

Campaigned?
All

College
All

Department
All

3274
Total Students

Sam ID	College	Department	Good Standing (RF)	Likelihood of Good Standing (RF)	Good Standing (LM)	Predicted GPA (LM)
[redacted]	Humanities and Social Sciences	Political Science	Y	61.7%	N	1.74
	Arts and Media	Art	Y	50.5%	Y	2.09
	Science and Engineering Tech	Agricultural Sciences	Y	54.4%	N	1.85
	Science and Engineering Tech	Biological Sciences	Y	67.3%	N	1.96
	Criminal Justice	Criminal Justice & Criminology	N	39.6%	N	0.97
	Science and Engineering Tech	Biological Sciences	N	34.1%	N	0.99
	Business Administration	Finance & Banking	N	31.8%	N	1.15
	Science and Engineering Tech	Agricultural Sciences	N	43.3%	N	1.71

Results

Model-Identified FTF (Fall 2024)

Week 4	Week 7	Total Unique FTF
297	251	366

Other Identified FTF from Alerts and ASC Campaigns: 775

1141

Total unique FTF intervened by the ASC thus far
(including model-identified students)

80 of the 297 students identified by the models in week four were later reported by their faculty in progress reports in week seven.

Flipping the Script

- Aggregating results by:
  - Course section
  - Instructor
- Coupling results with:
  - Historical DFQ rates
  - Instructor evaluation scores
- Identifying "**toxic combinations**" - sections with:
  - Sections with a high density of at-risk students
  - For a course with a high historical DFQ rate and
  - An inexperienced or low-rated instructor
- Using results to inform **instructor scheduling** and to offer pre-term **faculty development** and wraparound **academic support services** (e.g. embedded tutoring or supplemental instruction)

Data Analytics and Decision Support

The Pivot

Predicting Pass/Fall: Unweight Grade vs Overall grade

Overall Grade

- ⟶ Calculated (weighted) grade field
- ⟶ Requires additional setup by instructor
- ⟶ Low utilization, few observations
- ⟶ Performs slightly better than alternative when predicting Pass/Fail (80%)
- ⟶ Significantly outperforms alternative when predicting end term grade

Average Normalized Grade

- ⟶ Unweighted average of all a student's grades
- ⟶ Does not require additional setup by instructor
- ⟶ Almost double the observations (some still missing due to instructors inputting no grades)
- ⟶ When predicting pass vs fail, performs almost as well as Overall Grade (78%)
- ⟶ Underperforms compared to Overall Grade when predicting end term grade

Fall 2025 Avg Normalized Score vs Overall Grade

Date	Unweighted Grade Accuracy	Observations	R2	Overall Grade Accuracy	Observations	R2
Sep 15	77%	10,650	.08	78%	5,202	.16
Oct 1	78%	12,117	.15	80%	7,224	.24
Oct 31	80%	12,775	.23	82%	8,225	.30
Dec 1	81%	12,907	.26	83%	8,969	.33

Generating EAM

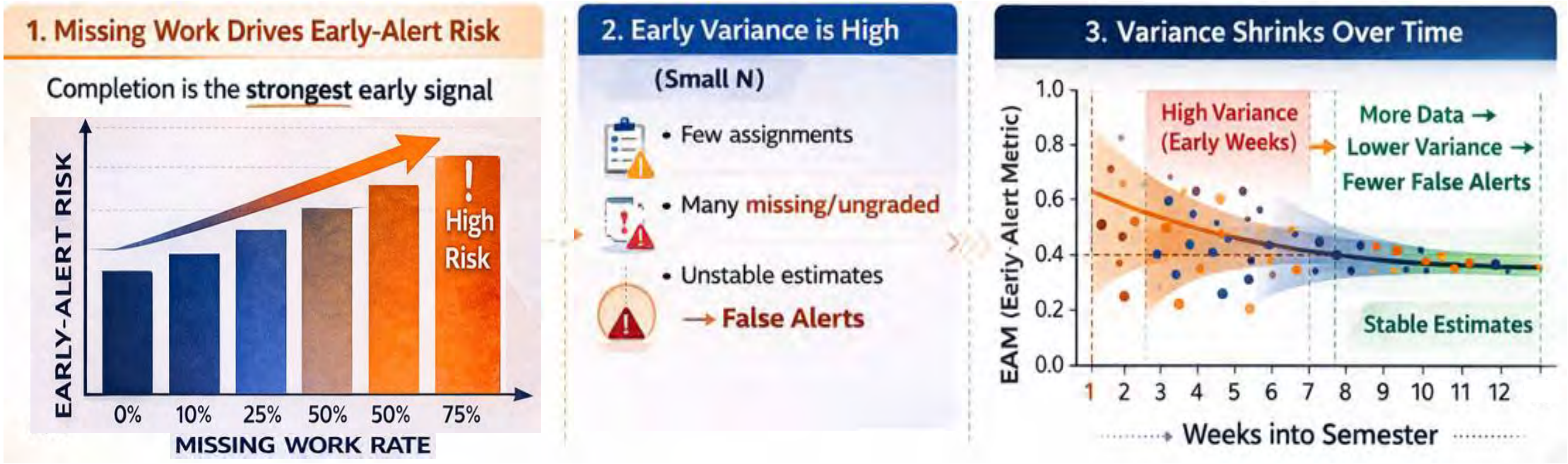
Combine the observation count of unweighted grade with the boosted predictive power of overall grade by swapping in the overall grade where available.

Resulting measure is our EAM

```
COALESCE(r.overall_normalized_score, r.avg_normalized_score)
* (r.assignments_completed_in_window / NULLIF(s.assignments_released_in_window, 0))
```

The EAM Methodology - Problem

Problem: How do we estimate early-alerts without making false positives and false negatives?



The EAM Methodology – Proposed Estimator

Engagement $(C_i) = \dfrac{\textit{No of assignments completed and graded}}{\textit{Total no of assignments released}}$

Performance $(P_i) = \dfrac{\sum_i^n \textit{Normalized score of the } i^{th} \textit{ assignment}}{\textit{No of assignments graded } (n)}$

EAM_i = **Engagement** (C_i) (Completion Rate) × **Performance** (P_i) (Average Normalized Score)

Higher the EAM, lower the risk

Individual Test Performance (Raw Scores 1–5)

i^{th} student	Test 1	Test 2	Test 3	Test 4	Test 5
Student A	Test 1: 60	Missing	Missing	Missing	Missing
Student B	Test 1: 20	Test 2: 20	Test 3: 20	Missing	Missing
Student C	Test 1: 70	Test 2: 60	Test 3: 65	Test 4: 60	Missing
Student D	Missing	Missing	Missing	Missing	Missing

i^{th} student	Completion Rate (✓)	Avg. Normalized	EAM (EAM*)	Percentile Rank
Student A	0.2 (20%)	0.6 (60%)	EAM* 0.12 Low	2
Student B	0.6 (60%)	0.2 (20%)	EAM* 0.12 Low	2 ⚠
Student C	0.8 (80%)	0.6375 (64%)	EAM* 0.51 Moderate	1
Student D	0.0 (0%)	0.0 (0%)	EAM* 0.00 0 (Low)	4

Missing work is defined as either work the student has not completed or work the instructor has not yet graded.

The EAM Methodology – Two Competing Estimators

Sensitive: Multiplicative Model

$$EAM_i = \text{Engagement } (C_i) \text{ Completion Rate} \times \text{Performance } (P_i) \text{ Average Normalized Score}$$

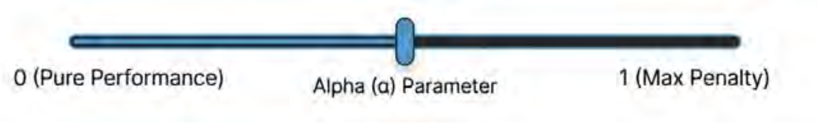
- Asymptotically unbiased
- Biased for early alerts
- High early variance
- More likely to produce ties

Unnecessary alerts

Conservative: Additive Model

$$Alt\_EAM_i = \alpha \times C_i + (1- \alpha) \times P_i \mathbb{I}(C_i > 0)$$

$\alpha \in [0,1]$: *penalty parameter for missing work*



- Understate average risk (biased downward)
- Biased for early alerts
- Low early variance
- Less score clumping

Missing at-risk students

The EAM Methodology – Stabilized EAM (Shrinkage Estimator)

When an estimate is noisy, it is optimal to shrink it toward a stable reference value.

$$\textbf{Stabilized } EAM_i = \mathbb{E}[EAM_i | data]$$

$$\textbf{Stabilized } EAM_i = \begin{cases} C_i \times EAM_i + (1 - C_i) \times \overline{EAM_S}, & C_i > 0 \\ 0, & C_i = 0 \end{cases}$$

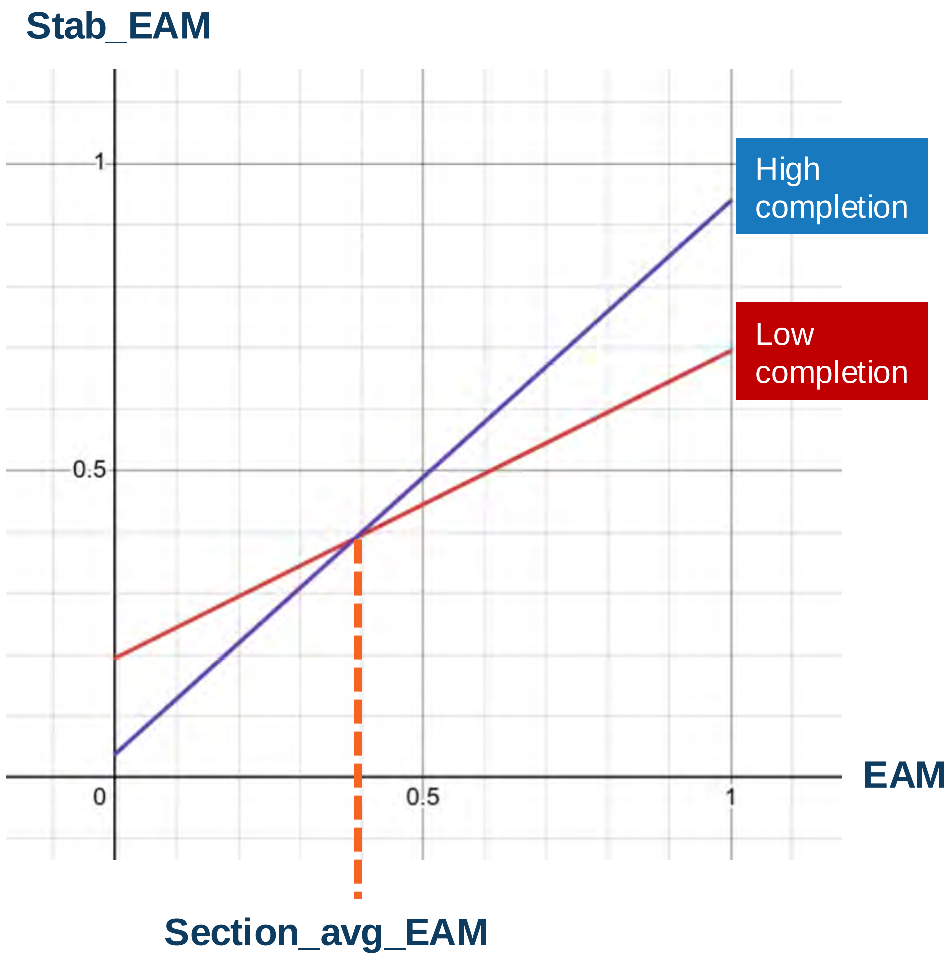
$$\overline{EAM_S} = \frac{\sum_{i=1}^{n} EAM_i}{n} \text{ or } \frac{\sum_{i=1}^{n} Alt\_EAM_i}{n}$$

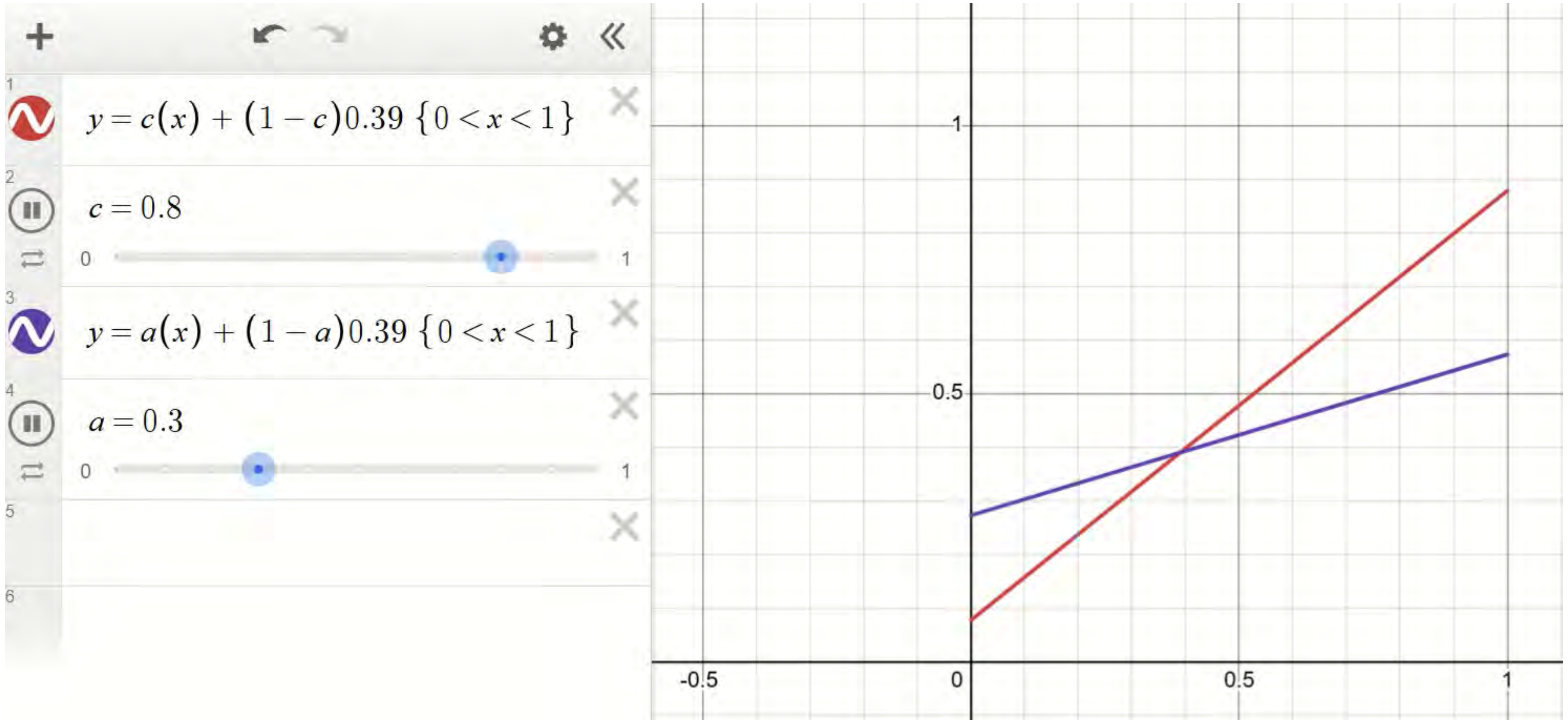
$\overline{EAM_S}$: Section average EAM

n: *Total number of current enrollments of the section*

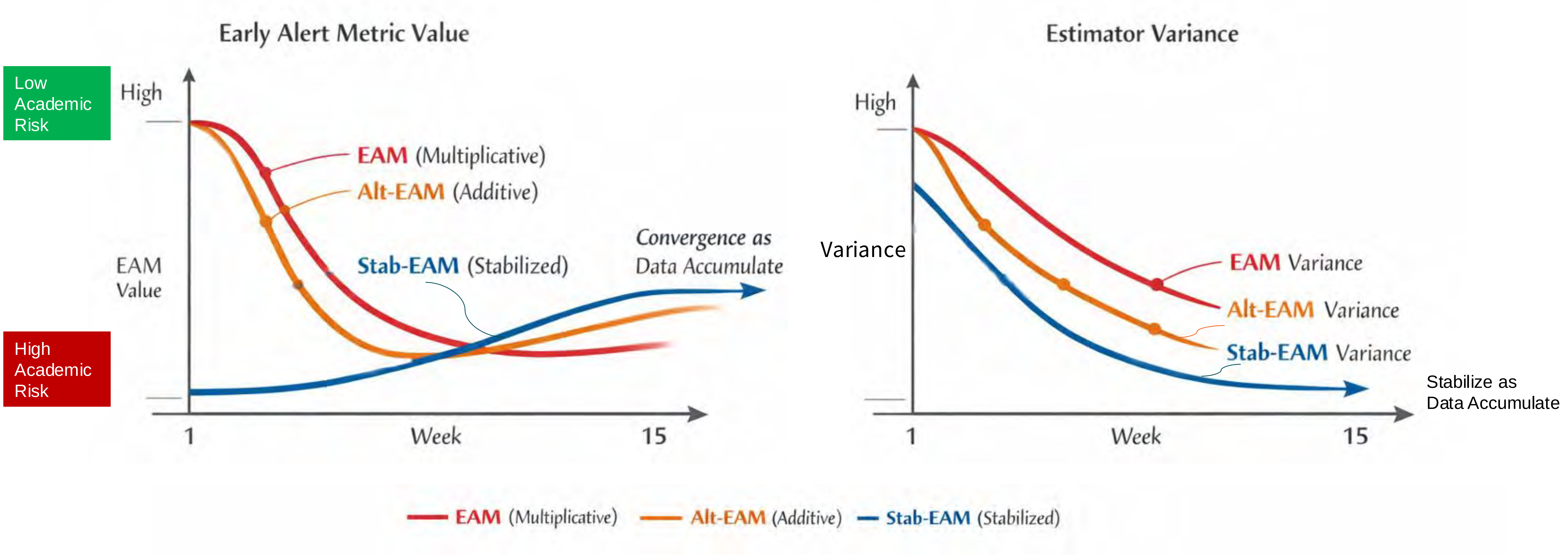
Student	Test 1	Test 2	Test 3	Test 4	Test 5	C_i	P_i	EAM_i	$Stble\_EAM_i$	Percentile Rank based on Stab_EAM
A	60	.	.	.	.	0.2	0.6	0.12	**0.174**	**2**
B	20	20	20	.	.	0.6	0.2	0.12	**0.147**	**3**
C	70	60	65	60	.	0.8	0.6375	0.51	**0.4455**	**1**
D	.	.	.	.	.	0.0	0.0	**0.0**	0.1875	**4**
							$\overline{EAM_S}$	**0.1875**		

The EAM Methodology – Interpreting Early Risk Signals





The EAM Methodology – Estimator performance



Data Analytics and Decision Support

The Tool

Turning Alert Predictions into Faculty-led Action

- A faculty-facing tool uses this early-alert model and student performance data to close the gap
- Faculty can easily monitor student progress in real-time and identify early indicators of academic risks
- Early identification of at-risk students can lead to faster outreach, early interventions and better support

Faculty-Facing Tool: Instructor Profile

- **Instructor-Level Insights**
  - Compare personal course DFQ rates to course average DFQ rates
  - Identify courses with higher risks
  - Understand how their outcomes compare to faculty peers

Faculty-Facing Tool: Course Performance View

- **Course-Level Insights**
  - View real-time student performance indicators in their own courses
  - Identify students showing early signs of academic risks
  - Monitor alerts, assignment completion rates, student percentile rankings

16.50	71.6%	66.4%
Avg. Assignments Completed	*Avg. Unweighted Average*	*Avg. Weighted Grade*

Course Performance View

Step 1: Select the semester view

Standard View | **Assignment Completion Flexibility**

Step 2: Select the desired filters

Course

Spring 2026 ENGL 2555 02 Dramatic E…

Student Name: All

Care Unit Responsible: All

Support Needed: All

Student	Support Needed	Care Unit Responsible	Student's Graded Assignments	Unique Graded Assignments in Grade Book	Unweighted Average	Weighted Grade	Early Alert Recommended?	Class Section Rank
Toby McSquiggles	Low Support	First Gen Center	17	17	80.1%	68.8%	No	38%
Beatrice Bumblepants	Low Support	ASC	18	17	82.4%	75.2%	No	46%
Wally Snickerdoodle	Low Support	TRIO	17	17	80.9%	72.1%	No	54%
Bongo McFlapjack	Low Support	Veteran	17	17	77.3%	73.6%	No	62%
Patty O'Pancake	Low Support	ASC	17	17	76.1%	77.5%	No	77%
John Doe	Low Support	Athletics	17	17	78.6%	81.0%	No	85%
Chester Von Quack	Low Support	ASC	17	17	83.3%	85.9%	No	100%
Marge Picklewhistle	High Support	ASC	17	17	41.2%	25.0%	Yes	0%
Doris Fluffernugget	High Support	Athletics	14	17	49.0%	39.6%	Yes	8%
Lenny Fizzlebop	Moderate Support	ASC	16	17	67.5%	65.2%	Yes	15%

Faculty-Facing Tool: Student Progress View

- **Individual-Student Insights**
  - See alerts across a student's other courses
  - Identify student's performance across their academic courseload
  - Distinguish between a course-level risk or overall academic risk

Student Progress View

Student Name

John Doe

Student Email

joh218@shsu.fake.email.edu

Care Unit Responsible

Athletics

1

Number of Recommended Early Alerts

Course	Support Needed	Early Alert Recommended?
Spring 2026 ARTS 1366 04 Doodling as Visual Communication	Low Support	No
Spring 2026 BIOL 2766 03 Sleepy Animals & Nap Behavior	Low Support	No
Spring 2026 ENGL 2555 02 Dramatic Email Writing	Low Support	No
Spring 2026 ENGL 2555 02 Fashion That Shaped Society	Low Support	No
Spring 2026 MATH 3310 02 Statistics for Predicting the Weather	Moderate Support	Yes

Faculty-Facing Tool: Assignment Completion Flexibility

File Export Share Explore Subscribe Set alert Monitor Edit ··· Copilot

Assignment Completion Importance

0.5

Student	Support Needed	Care Unit Responsible	Student's Graded Assignments	Unique Graded Assignments in Grade Book	Unweighted Average	Weighted Grade	Early Alert Recommended?	Class Section Rank
Beatrice Bumblepants	Low Support	ASC	17	17	82.4%	75.2%	No	89%
Bongo McFlapjack	Low Support	Veteran	14	17	77.3%	73.6%	No	44%
Chester Von Quack	Low Support	ASC	13	17	83.3%	85.9%	No	100%
John Doe	Low Support	Athletics	17	17	78.6%	81.0%	No	56%
Patty O'Pancake	Low Support	ASC	17	17	76.1%	77.5%	No	33%
Toby McSquiggles	Low Support	First Gen Center	17	17	80.1%	68.8%	No	67%
Wally Snickerdoodle	Low Support	TRIO	15	17	80.9%	72.1%	No	78%
Doris Fluffernugget	High Support	Athletics	14	17	49.0%	39.6%	Yes	11%
Lenny Fizzlebop	Moderate Support	ASC	8	17	67.5%	65.2%	Yes	22%
Marge Picklewhistle	High Support	ASC	6	17	31.2%	29.1%	Yes	0%

Filters

132%

Future Directions

batchuid_person	person_name	course_identification	grade	normaliz...	row_inserted_time
		MATH1420	C	.7620253	02sep2025 03:52:49
		MATH1420	C	.7837417	04sep2025 03:57:14
		MATH1420	C	.7837417	05sep2025 03:54:51
		MATH1420	C	.7208763	06sep2025 03:59:49
		MATH1420	C	.7368773	07sep2025 03:55:34
		MATH1420	C	.7806607	08sep2025 03:52:07
		MATH1420	C	.7934616	09sep2025 03:56:42
		MATH1420	D	.6813895	10sep2025 03:58:46
		MATH1420	C	.7120097	11sep2025 03:56:52
		MATH1420	C	.7946839	12sep2025 03:57:36
		MATH1420	B	.8090529	16sep2025 03:57:47
		MATH1420	B	.8058128	17sep2025 03:58:04
		MATH1420	B	.8087994	19sep2025 03:59:08
		MATH1420	C	.7981516	20sep2025 03:55:21
		MATH1420	C	.7981516	21sep2025 03:57:04
		MATH1420	C	.7981516	22sep2025 03:57:36
		MATH1420	C	.7548649	23sep2025 03:56:22
		MATH1420	C	.7494025	25sep2025 03:59:02
		MATH1420	C	.7570879	26sep2025 03:56:20
		MATH1420	C	.7627142	27sep2025 03:56:25
		MATH1420	C	.7627142	28sep2025 03:56:18
		MATH1420	C	.7627142	29sep2025 03:53:01
		MATH1420	C	.7648996	30sep2025 03:56:34
		MATH1420	C	.7648996	01oct2025 03:58:17
		MATH1420	C	.7477505	03oct2025 03:55:05
		MATH1420	C	.7483686	04oct2025 03:55:59

Logistic regression

Number of obs = 12,117
LR chi2(8) = 3138.63
Prob > chi2 = 0.0000
Pseudo R2 = 0.2274

Log likelihood = -5331.7839

pass	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
stabilized_eam	5.812657	.1546064	37.60	0.000	5.509634	6.11568
dfq_rate_course	-2.946598	.2845579	-10.36	0.000	-3.504322	-2.388875
dfq_rate_instructor	-1.399725	.2387942	-5.86	0.000	-1.867753	-.9316972
black	-.3115037	.0542324	-5.74	0.000	-.4177972	-.2052101
high_school_gpa	.795037	.055141	14.42	0.000	.6869626	.9031115
sat_concordance	.0007593	.0002816	2.70	0.007	.0002074	.0013113
first_gen	-.0415339	.0499529	-0.83	0.406	-.1394397	.0563719
male	-.1179621	.0507187	-2.33	0.020	-.2173689	-.0185554
_cons	-5.175967	.3297845	-15.69	0.000	-5.822332	-4.529601

Logistic regression

Number of obs = 7,224
LR chi2(8) = 2364.56
Prob > chi2 = 0.0000
Log likelihood = -3156.9048
Pseudo R2 = 0.2725

pass	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
overall_normalized_score	5.410308	.1562157	34.63	0.000	5.104131	5.716485
dfq_rate_course	-3.174876	.3530827	-8.99	0.000	-3.866906	-2.482847
dfq_rate_instructor	.0288644	.3029069	0.10	0.924	-.5648221	.6225509
black	-.3160967	.0707287	-4.47	0.000	-.4547223	-.177471
high_school_gpa	.7547085	.0725044	10.41	0.000	.6126026	.8968145
sat_concordance	.0000428	.0003721	0.12	0.908	-.0006865	.0007721
first_gen	-.1517699	.0653329	-2.32	0.020	-.2798201	-.0237197
male	-.0296006	.0663336	-0.45	0.655	-.1596121	.100411
_cons	-4.576842	.4257375	-10.75	0.000	-5.411273	-3.742412

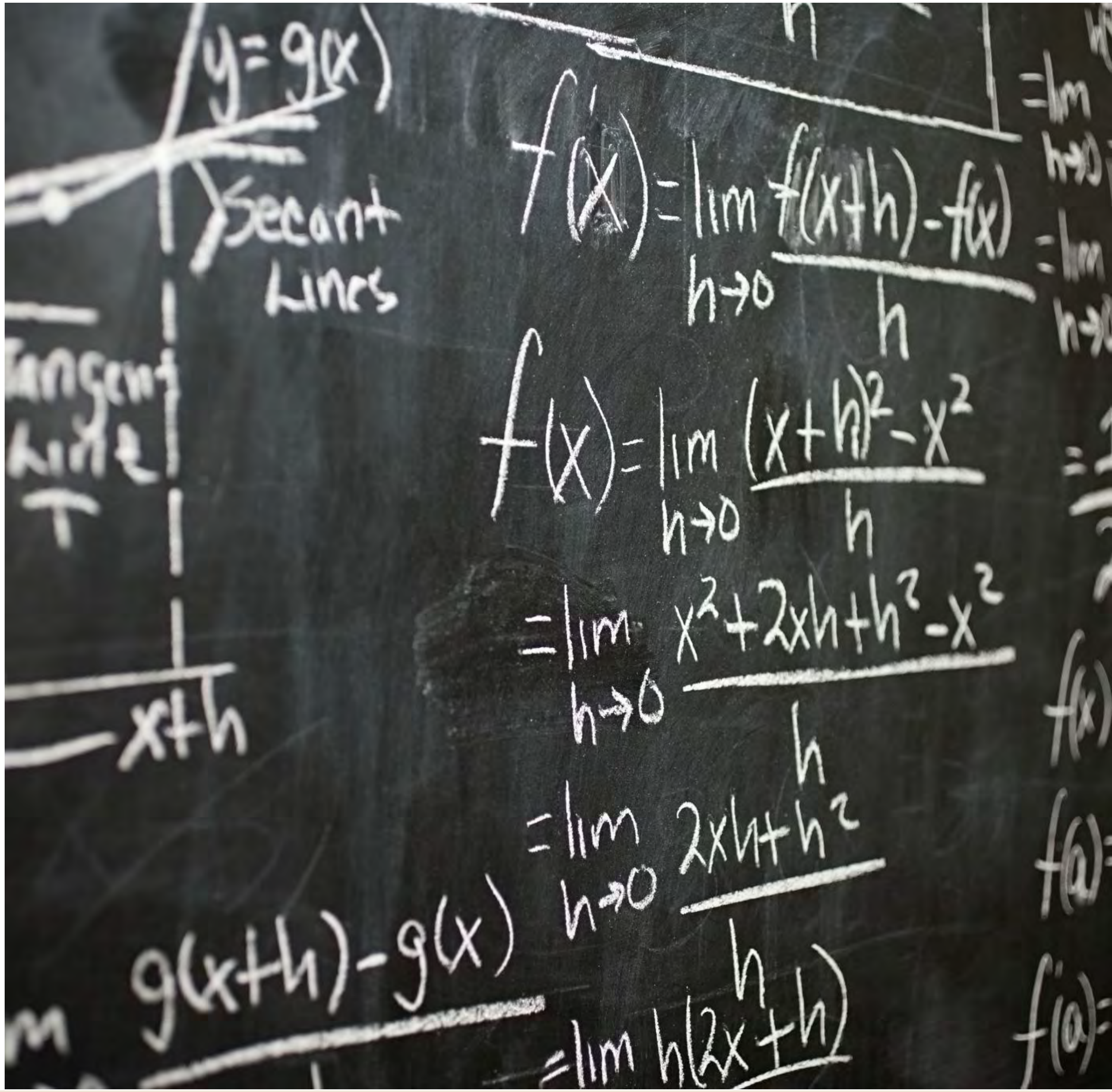
Fall 2025 Full Model Results

Date	EAM Accuracy (no grade)	Observations	R2	EAM Accuracy (w/ grade)	Observations	R2	Overall Grade Accuracy	Observations	R2
Sep 15	79%	5,448	.17	79%	5,202	.18	78%	5,202	.20
Oct 1	80%	4,893	.20	80%	7,224	.24	80%	7,224	.27
Oct 31	82%	4,550	.25	82%	8,225	.28	82%	8,225	.32
Nov 1	82%	3,938	.27	83%	8,969	.28	83%	8,969	.35

- Results are a slight improvement to running just the grade columns on their own

Benefits of using Blackboard Grades

- Simple and compact with strong predictive power
- Easy to understand & easy to explain
- Quick & easy to run
- Course agnostic
- Student type agnostic



Q&A and Discussion

Scan the QR code to complete the session survey.

2026