Identifying SRS in Math With High School Transcript Data

Executive Summary

This brief summarizes the findings from an analysis looking at how information contained in students' high school transcripts can be used to identify math as a reason a student is flagged as a Student Recommended for Support (SRS). Historically, students have mostly been flagged as SRS because of math based on ACT/SAT scores determined by state level criteria (at or below 490 for the SAT or 18 for the ACT). Students can also be flagged as SRS because of math based on Admission's holistic review, but there are currently no set standards for determining this math status.

However, beginning with the FA21 cohort, CSU adopted a test optional policy and fewer students have chosen to submit their standardized test scores when applying. As a result, the traditional process used to identify students in need of additional support in math while at CSU cannot be relied upon in the same way. Instead, admissions staff will need to consult other sources of information about students' prior academic experience in math, and high school transcripts offer the best alternative.

Using high school transcript data compiled by CSU's Office of Admissions for a random sample of students, IRPE set out to explore what transcript indicators – if any – could be used to reliably identify whether a student should be flagged as SRS in math. The intent is to establish a concrete definition for "struggled with math in high school" among the sample of students. To develop this definition, we use machine learning and logistic regression analyses to explore the predictive power of high school transcript data on first-year math success in college.

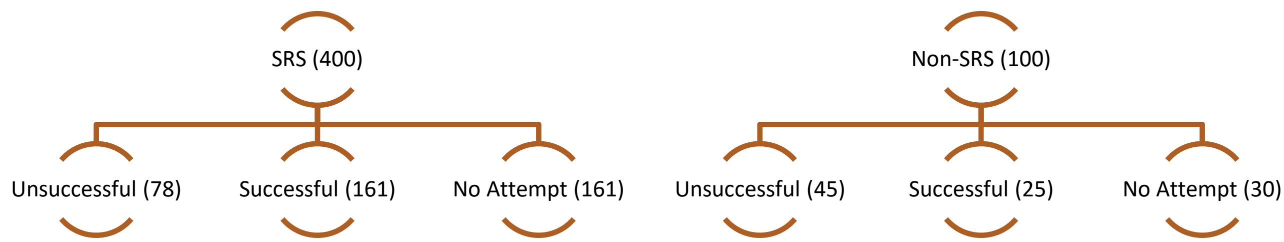
We find that out of 11 high school transcript dimensions, only one is meaningfully able to predict whether a student will be unsuccessful or successful in math in their first year: students who earn 1 or more Ds or Fs total in grades 9th through 12th (counted on a semester basis) are more likely to have an unsuccessful math completion than their otherwise similar peers who also attempt math. Importantly, we were not able to find any transcript indicators that could reliably predict whether a student would attempt math or not in their first year.

Still, though we are able to identify the number of Ds/Fs earned in high school as a important predictor of first-year math success, using this as the single decision criteria for SRS status would still be problematic because we would expect to misclassify students as SRS or non-SRS about a third of the time. Our findings are limited by our relatively small sample size, and future samples drawn for analyses may need to be larger than 500 students to yield more conclusive findings. However, this analysis does help frame future investigations because it highlights a key transcript variable to focus on in more detail going forward: the counts of unsuccessful math attempts in high school.

Data and Methods

Originally, IRPE pulled a stratified random sample of 500 full-time, first-time undergraduates in the FA19 cohort. Students with prior learning assessments in math were excluded, though students with AP/IB/DE credits were included. The sample was pulled such that SRS students overrepresented non-SRS students 4-to-1, and it included students who were successful in math their first year, unsuccessful in math their first year, and those who did not attempt math at all in their first year.[1] Figure 1 below shows the original sample headcounts by first-year math status, separately for SRS and non-SRS students.

Figure 1. Headcounts () of students in stratified random sample by SRS and Non-SRS



After the sample had been identified, Admissions reviewed the high school transcripts of these 500 individuals and appended additional data about their high school math experiences believed to be informative about their first-year college math experiences. Admissions staff added data on 11 dimensions in total including how many unique math courses students took while in high school, whether they took AP/IB/DE courses, counts of Ds and Fs by semester, etc. Appendix Table 1A defines these 11 dimensions in more detail.

With this newly appended data, IRPE used statistical analyses to examine the relationship between high school transcript math indicators and students' first-year math attempt. Because this analysis is mostly exploratory and not driven by a specific hypothesis about which transcript indicators predict first-year math success, we use a machine learning approach to explore patterns in the data and supplement with logistic regression analysis.

Findings

Summary Statistics

Before performing any statistical analyses, we compared demographics and high school transcript information between the three first-year math outcomes: 1) Unsuccessful, 2) Successful, and 3) No Attempt. Importantly, though the original random sample contained 500 students, transcript data could only be collected for 481 students.[2] This is the headcount we restrict our focus to in all statistical analyses going forward.

Table 1 displays the demographic and transcript means by first-year math status. Bolded numbers indicate that these are means are statistically significantly different from the Unsuccessful first-year math mean through a formal t-test (p-level of 0.05).

[1] Students were not double counted, therefore, if a student is labeled "Successful", they only had successful math attempts in their first year. Audits and incompletes of math courses were excluded from the sample as well because they are not considered successful nor unsuccessful math attempts.

[2] The 19 students without transcript data are those with international credentials who are not evaluated by Admissions, but rather the Office of International Programs. The composition of the sample with 481 students is similar to that of the original 500 in terms of the relative share who are SRS vs. non-SRS and those who first-year math was unsuccessful, successful, or not attempted.

Table 1. Summary Statistics by First-Year Math Status

	Unsuccessful	Successful	No Attempt
N	118	174	189
Demographics			
SRS	62.7%	**85.6%**	**84.1%**
Female	55.9%	58.6%	59.8%
First Generation	28.0%	32.2%	**39.2%**
Racially Minoritized	37.3%	31.6%	41.3%
Pell	30.5%	28.7%	37.6%
High School Transcript			
Count of unique math courses	4.26	4.15	4.28
Student had non-college prep courses	8.5%	6.3%	11.1%
Student had extended/bridge year	2.5%	1.7%	1.6%
Student had math in final year	94.9%	90.2%	**89.9%**
Student is at/above Alg2 in fourth year	92.4%	93.7%	88.8%
Student took AP, IB, and/or DE courses	32.2%	26.4%	27.5%
Student went directly from high school to college	93.2%	93.1%	92.6%
Count of Ds/Fs in Semester	1.09	**0.41**	**0.72**
Count of Ds/Fs in Quarter	0.00	0.05	0.06
Count of retakes when earning a C or higher	0.13	**0.05**	0.10
Count of retakes when earning less than a C	0.00	0.01	0.02

Notes: Proportions are displayed as shares (%) while continuous variables (i.e., counts) are displayed as decimals.

For example, while 28% of those students Unsuccessful in math are first generation, 32.2% of Successful students are first generation, and 39.2% of students who made No Attempt in math are first generation. Only the 39.2% for the No Attempt group is statistically different from 28% observed for the Unsuccessful group, which is why it is bolded.

There are three main takeaways to get from looking at Table 1. The first concerns mean differences in the high school transcript data. Looking at the 11 high school transcript dimensions, we see that there are several places where the means vary meaningfully from the Unsuccessful group mean. Although, there is only one dimension where both the Successful group mean and the No Attempt group mean are statistically significantly different from the Unsuccessful group mean: the count of Ds or Fs earned on a semester basis from 9th grade through 12th grade.

The second takeaway concerns consideration of how the sample was constructed and how this affects our interpretation of the machine learning and logistic regression analyses. This is best illustrated by the differences in the shares of students flagged as SRS in FA19. While 62.7% of students who were Unsuccessful in math their first year were flagged as SRS, 85.6% of Successful students and 84.1% of No Attempt students were flagged as SRS. The bolded means for Successful and No Attempt groups mean that they are statistically significantly different from that of the Unsuccessful group.

Intuitively, these mean differences may seem strange; the share of SRS students is notably lower among those with unsuccessful first-year math attempts than the other two groups. However, this data point is mostly an artifact of how the sample was constructed. As Figure 1 shows, out the 400 SRS students pulled, 161 (40.3%) were successful in math, 161 (40.3%) did not attempt math, and 78 (19.5%) were unsuccessful in math. As a result, the full sample of students is skewed towards having a higher relative share of SRS students who either were successful in math or did not attempt it at all in their first year.

The third and final takeaway has to do with the small means for some transcript variables and therefore their reduced ability to predict first-year math outcomes for students. For instance, the mean count of Ds or Fs earned on a quarter basis from 9th through 12th grade is extremely small across all three groups at 0.0, 0.05, and 0.06, respectively. In addition, the relative share of students who had an extended/bridge year is also quite small across all three groups at less than 3%. These small numbers hint at what we observe in more detail below: for some variables, not enough students are flagged with that transcript criteria for us to observe a meaningful relationship between them and whether a student attempted math and/or whether they were successful in their attempt.

Machine Learning

The first analysis we perform is a machine learning technique called categorical and regression tree analysis (CART). CART is an analytic technique that helps determine the most “important” variables in a particular dataset based on their ability to predict the outcome of interest. In our analysis, the outcome we are predicting is students’ first-year math outcome: 1) Unsuccessful, 2) Successful, or 3) No Attempt. The primary variables we consider for importance are the 11 high school transcript indicators.[3] The final product from the CART analysis, as the name implies, is a decision tree.

The first step in CART involves splitting our dataset of 481[4] observations into a “training” dataset and a “testing” dataset. The training dataset is the one that the model uses to evaluate the importance of different high school transcript indicators and to construct a decision tree. The testing (sometimes called “evaluation”) dataset is the one where we can “test” how well the decision tree works on new, unseen data. Typically, the training dataset is comprised of about 70% of the observations, while the other 30% are in the testing dataset. In this case, that meant 336 students were placed in the training dataset and 143 were placed in the testing dataset.

However, when we ran the CART analysis for the full sample of students including all three math outcomes (Unsuccessful, Successful, and No Attempt) our decision tree had very low predictive power (~29%). In fact, the final decision tree did not attempt to classify how many students would have No Attempt in first-year math. For this reason, we chose to eliminate this group to focus on just the 1) Unsuccessful (N=118) and 2) Successful (N=174) groups of students. Now, our testing dataset contains 205 students and our testing dataset contains 87 students out of the total 292 who attempted math.

Figure 2 shows the resulting decision tree from the CART analysis using the training dataset. Altogether, this decision tree had predictive accuracy of 69%, meaning that if we follow the logic of the tree for every student in the training dataset (205), we would have classified their first-year math status (unsuccessful or successful) correctly 69% of the time.

Figure 2. Decision tree for first-year math status, among those who attempted math

[3] We also include demographic characteristics (e.g., female, Pell status, first generation, etc.) in additional models, but these have no meaningful relationship with first-year math outcomes in this data.
[4] Really, we only performed this analysis on 479 observations because 2 of the 481 students were missing information about whether they were at or above Algebra 2 in their fourth year and were therefore automatically dropped from the model.

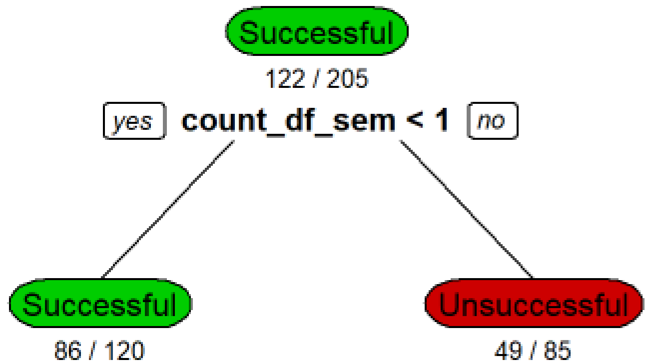


Figure 2 shows several things. First, the green oval at the top tells us that 122 of the 205 students (59.5%) in the training data were actually successful in their first-year math attempt. The second thing it tells us is that out of the 11 transcript dimensions it considered, only one was found to be an important predictor of students' first-year math status. That dimension was the count of how many Ds or Fs a student had over a semester timeline across their 9th through 12th grades. The third thing it tells us is that if a student had less than 1 (meaning 0) Ds or Fs, the student is predicted to have a successful first-year math experience. If the student has 1 or more Ds or Fs on their transcript, they are predicted to have an unsuccessful first-year math experience.

The fractions below the bottom "leaves" of the decision tree show how well the tree predicted each outcome specifically. For example, the decision tree predicted that 120 students would be successful in math their first year *just using the count of Ds and Fs as its criteria*, when in reality, only 86 of these 120 students were successful in math. The other 34 students (i.e., 120 minus 86) were misclassified as successful by the tree; in reality, they were unsuccessful in math.

Relatedly, the decision tree predicted that 85 students would be unsuccessful in math their first year *just using the counts of Ds and Fs as its criteria*, when in reality only 49 of these 85 students were unsuccessful. The other 36 students (i.e., 85 minus 49) were misclassified as unsuccessful by the tree; in reality, they were successful in math. Note that the tree was better at predicting math success (86/120=71.7%) than math unsuccess (49/85=57.6%) when counting students total Ds and Fs earned in high school as its only decision criteria.

Figure 3 shows how well this decision tree held up among the other 87 observations in the testing/evaluation dataset as a "heatmap". The "true" outcome of a student is on the horizontal axis while the predicted outcome using just the count of Ds/Fs – and therefore whether it exceeds 0 – is on the vertical axis.

Figure 3. Heatmap of decision tree accuracy for first-year math status, among those who attempted math

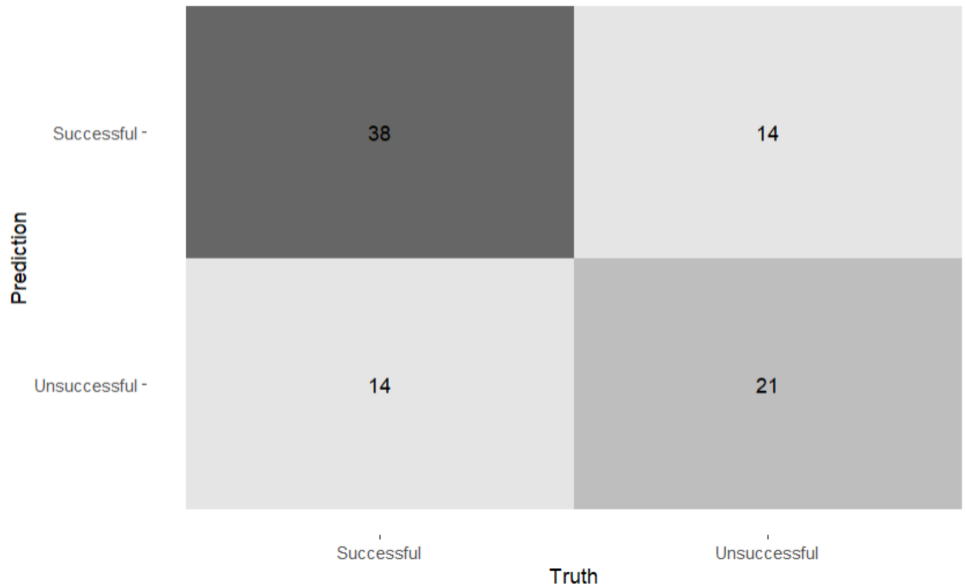


Figure 3 shows that a total of 52 of the 87 students were successful in math (i.e., 38 plus 14), while the remaining 35 (i.e., 14 +21) were unsuccessful in math in "truth" (the horizontal axis). It also shows that it predicted correctly that 38 of the 52 Successful students succeeded (73.1%) and that 21 of the 35 Unsuccessful students did not succeed (60%). As was true for the training dataset, the decision tree did a better job accurately predicting successful math completion than unsuccessful math completion when using just Ds and Fs as a sorting criteria (hence, the darker shade of the 38 box).

From the CART analysis, the only meaningful predictor of first-year math success taken from the high school transcript data was the count of total Ds and Fs earned in 9^{th} through 12^{th} grade (summed on a semester, as opposed to a quarter, basis). However, even then, we can see that the decision tree still has a relatively high misclassification rate at about 32% for both the training and the testing datasets. This suggests that even though the count of Ds and Fs may be the most meaningful indicator of first-year math success, it is not a perfect, nor even incredibly reliably indicator.

Logistic Regression

We also supplemented the CART analysis with a more traditional logistic regression analysis as well. Because our outcome of interest – student's first-year math status – can take on three values, each one dependent on whether a student attempts math at all, we split the regression into two parts. First, we explore what factors would predict whether a student would attempt math at all in their first year, and then we explore what factors predict unsuccessful math completion among those students who did attempt math. Table 2 shows the odds ratios from logistic regressions predicting first-year math status.

Table 2. Odds ratios from logistic regressions predicting first-year math status

	Attempted Math		Unsuccessful Math	
VARIABLES	1	2	3	4
Student took less than 4 unique math courses	1.914		0.456	
	(0.830)		(0.282)	
Student took more than 4 unique math courses	1.016		0.939	
	(0.235)		(0.296)	
Student had non-college prep courses	0.756		0.954	
	(0.278)		(0.525)	
Student had math in final year	1.569		1.556	
	(0.638)		(0.888)	
Student is at/above Alg2 in fourth year	1.745		0.472	
	(0.776)		(0.312)	
Student took AP, IB, and/or DE courses	1.089		1.482	
	(0.240)		(0.416)	
Student went directly from high school to college	0.907		1.074	
	(0.354)		(0.577)	
Student retook a D/F course	0.647		0.984	
	(0.264)		(0.542)	
Female	1.030	0.998	0.973	0.997
	(0.208)	(0.195)	(0.255)	(0.259)
First Generation	0.756	0.808	0.808	0.850
	(0.175)	(0.184)	(0.256)	(0.269)
Racially Minoritized	0.861	0.846	1.234	1.307
	(0.187)	(0.179)	(0.357)	(0.371)
Pell Recipient	0.810	0.802	1.029	1.047
	(0.187)	(0.179)	(0.332)	(0.324)
Student earned 1 or more Ds/Fs in a semester	1.215	1.138	3.643***	3.539***
	(0.263)	(0.222)	(1.009)	(0.901)
Observations	479	481	292	292
Model Chi2	13.50	6.094	32.16	26.78
Model Degrees Freedom	13	5	13	5
Pseudo R2	0.0209	0.00950	0.0872	0.0720

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Columns 1 and 2 show that no transcript variables, nor any demographic characteristics, have a statistically significant correlation with whether a student will attempt math or not in their first year at CSU.[5] However, when we restrict our focus to the 292 students who do attempt math in their first year in columns 3 and 4, we do observe a statistically significant relationship between the number of Ds and/or Fs a student earns (shaded rows). This relationship holds whether we control for just demographic characteristics or for the full suite of high school transcript variables.

Students who earned 1 or more Ds/Fs in a semester at any point between grades 9 through 12 have about 3.5 – 3.6 the odds of having an unsuccessful math experience in their first year compared to their peers with 0 Ds/Fs.

[5] Note that not all 11 transcript indicators are included in the logistic regressions because some are multicollinear (e.g., counts of Ds/Fs earned on semester and quarter basis) with one another, while others do not have enough power to detect a meaningful relationship (e.g., had an extended/bridge year).

Table 3 translates the odds ratios in column 3, the model with both transcript and demographic variables included, into predicted probabilities of unsuccessful math completion by number of Ds/Fs.

Table 3. 95% confidence intervals of predicted probabilities for unsuccessful math

	Unsuccessful Math		
Had 1+ Ds/Fs	42.7%	-	70.6%
Had 0 Ds/Fs	16.0%	-	36.8%

Notes: The predicted probability confidence interval assumes a female, non-first generation, non-minority, non-Pell student who took 4 unique math courses in high school, did not have non-college preparatory math, did take math their final year, was at the Algebra 2 level of above in their final year, did not have any AP, IB, or DE courses, and went directly to college from math. The prediction intervals tell us where we can expect to find a predicted probability of unsuccessful math based on the distribution of data in our sample.

The logistic regression predicts that a student with 1 or more Ds/Fs has predicted probability of unsuccessful math completion between 42.7% and 70.6% compared, while a peer identical along every other dimension in Table 4 – except for having 0 Ds/Fs – has a predicted probability of unsuccessful math completion between 16.0% and 36.8%. These confidence intervals are very wide, and while they do not overlap, it is important to note how imprecise these predicted probabilities are for both student types. For instance, a student with 1 or more Ds/Fs would still have a predicted probability of *success* in math of 29.4 – 57.3%. These probabilities are quite high. Relatedly, even the 16.0 – 36.8% probability range of *unsuccess* for students with 0 Ds/Fs is quite high. At the upper boundary, a student with 0 Ds/Fs still has predicted probability of not succeeding of 37%, only 6 percentage points less than the lower boundary of 42.7% for the students with 1 or more Ds/Fs.

Altogether, the primary takeaway from Tables 2 and 3 is that the logistic regression analysis corroborates what was found in the machine learning CART analysis: the count of Ds/Fs in a semester is the only meaningful predictor of first-year math success – among those who attempt math – for sample we use. However, as with the CART analysis, the logistic regression shows that if we were to classify students as SRS based purely on counts of Ds and Fs, there would still be substantial number of students likely to be misclassified.

Conclusion

This analysis set out to develop a concrete definition for “struggled with math in high school” using high school transcript indicators. Through comparison of summary statistics, machine learning techniques, and logistic regression, we find that the only meaningful predictor of student’s first-year math success is the number of Ds/Fs a student earns on a semester basis between grades 9 and 12. However, this predictor is only meaningful among students who attempt math at all in that first year. This analysis did not yield useful information to determine what high school transcript indicators are correlated with first-year math avoidance and we might need a larger sample to explore this phenomenon.

The decision tree that demonstrates the predicted relationship is relatively straightforward and simple, though we should note that we do not know exactly *which* math class students receive a D or F in while in high school and the model still has a high misclassification rate of around 32%. These are important limitations of the analysis. The findings of this brief apply specifically to this sample and additional samples may be needed to fully endorse a recommendation that 1 or more unsuccessful math attempts in high school should be the criteria used to flag a student as SRS because of math.

Appendix

Table 1A. High School Transcript Dimensions

1	Number of unique math courses
2	Student had any non-college math prep courses (i.e., earlier than Algebra 1) {no, yes}
3	Student had extended/bridge year (typically to bridge Algebra 2) {no, yes}
4	Student took math in their final year {no, yes}
5	Student was at or above Algebra 2 by fourth year (i.e., Pre-Calculus, Statistics, Calculus) {no, yes, ?=CBD}
6	Student took AP, IB, or DE courses {no, yes}
7	Student went directly from high school to college taking math {no, yes}
8	Number of Ds and/or Fs earned in math in a semester at any point between grades 9 through 12
9	Number of Ds and/or Fs earned in math in a quarter at any point between grades 9 through 12
10	Number of retakes of Ds or Fs when earning a C or higher in the math course
11	Number of retakes of Ds or Fs when earning less than a C in the math course