Calibration and Instructional Improvement: An investigation of course success rates in sequential STEM courses at Cosumnes River College

Research and Equity Office

Spring 2024

Author:

Paul Meinz

Brief Summary of Key Findings[1]

1. Two analyses established a negative association between an instructor's success rate in the prerequisite and the success rate of their students in the next level of a STEM sequence.
2. The data were not perfectly described by this expectable association. For example, some instructors had above median course success rate in both the prerequisite course and the subsequent course.
3. The aforementioned observations were used to build two useful operational definitions. Specifically, distinguishing between *instructional improvement* or *calibration* in sequential STEM courses should use *both* the success rate in the prerequisite course and the success in the next level.
    a. *Instructional improvement* may be indicated by an increase in success in the first course and equivalent or increased success in the next course (or vice versa).
    b. *Calibration* may be indicated by a trade-off between success in the first course and success in the next. For example, the success rate in the first course may increase accompanied by a decline in success in the next level.

Recommendations

1. Instructors within course sequences should consider using *both* success in the prerequisite course and success in the sequential course to assess instructional improvement or calibration.
2. Efforts to increase course success should emphasize and provide resources towards instructional improvement.
3. The Research Office at CRC should provide data visualization of sequential course success data to support instructors within sequences.
4. Future research should also focus on the joint probability of a student passing the first and second course in a sequence and the factors (instructional or otherwise) that maximize that probability.

Important Caveats

This study could not account for differences in the instructional effectiveness and standards of the instructor in the second level. This may impact the validity of the associations described in this report. Additionally, the findings reported here only apply to sequential courses. Distinguishing between calibration and instructional improvement in a single course is beyond the scope of this report.

[1] Reading these high-level findings is not a sufficient substitute for reading and digesting the entire report. Many findings are not summarized here.

Introduction

In Fall 2023, Cosumnes River College (CRC) began a broad discussion around improving course success rates – specifically focused on classroom instruction. These conversations have expectedly led to important questions around the instructional meaning of a course success rate. For example, how should one interpret high or low success rates and improvements/declines therein? Finding the answer (or answers) to this question seems essential to any conversation around improving or changing course success rates. It is this question that led to two broad research requests submitted by the Academic Senate at CRC in Fall 2023:

1. The first request sought to investigate the potential correlation between course success rates in sequential STEM courses. Specifically, do students in courses with lower success rates have higher success rates in the next level?
2. The second question sought to identify non-instructional factors related to course success rate (e.g. the time a course is offered, the term length, etc.). That is, do course features and student barriers lead to differing success rates?

In an effort to support the work of Academic Senate and the College, the Research Office sought to investigate these two extremely broad research questions. This report pertains specifically to the first question regarding STEM success rates and sequential courses. Here, the operational definition of a *success rate* is the percentage of A, B, C, or P grades out of the total number of enrollments in a course after the deadline for dropping a course.

The Meaning of High and Low Success Rates

The question that this report attempts to address – regarding success rates across sequential STEM courses – is belied by two frequently discussed alternative (although not mutually exclusive) interpretations of success rates[2]. One might believe or hypothesize that an above average success rate for a course means *effective instruction* – e.g. the instructor has the same standards as other instructors but they are more capable of helping students achieve those standards. On the other hand, one may believe or hypothesize that a below average success rate for a course means *higher standards*, e.g. the instructor is just as good as other instructors at teaching, but they have higher standards for content mastery. Put another way, changes in course success rate may reflect variability in effective instruction and/or variability in instructional standards. When looking at a course success rate, these two interpretations can be used alternatively/interchangeably as explanations for why a success rate is high or low. Indeed, it is not hard to see how success rate might reflect a mixture of the two. In practice, it is possible (at least theoretically) to have high standards and be effective at teaching to those standards (and vice versa).

[2] This is not intended to be an exhaustive list of the interpretations of success rates, but a simple observation that there are two frequently referenced interpretations.

Calibration vs. Instructional Improvement

These two confounded interpretations of a success rate lead to a related and important distinction. If an instructor works to improve success in their course, how should they interpret a change in course success rate? For example, an instructor may implement a change in their course and see an increase in course success. On the basis of the two interpretations, this change could mean more students learned the material (improvement in instruction), the course got easier without improving learning (lower standards), or a mixture of the two.

The investigation of sequential success in STEM courses provides an opportunity to address the distinction between *calibration* and *instructional improvement*. For the purposes of the present discussion, *calibration* is defined as a change in the standards of the course without a change in instruction. For example, one might increase the difficulty of their tests and assessments but not change the way information is conveyed to students. *Instructional Improvement* is defined here as an improvement in the capacity to effectively teach a subject. For example, one may come up with a particularly clever way of explaining a difficult concept – resulting in higher levels of student learning. Put succinctly, calibration holds instruction constant while changing standards and instructional improvement holds difficulty constant while changing instructional effectiveness. Although the question about sequential STEM success is important in of itself, any investigation of correlation in success between sequential STEM courses should work to build an initial operational definition of calibration and instructional improvement. Indeed, fully defining and measuring these two constructs could take a lifetime of research[3]. Even so, the primary goal of this investigation is to help provide an *initial framework* to distinguish between these two concepts in sequential courses of a STEM discipline. This in turn will support the discussions of Academic Senate and provide support to instructors hoping to improve how they teach.

The Present Investigation

To help build this framework, sequential course success was investigated in several large STEM pathways. Specifically, four pathways were selected on the basis of the number of students and instructors in those sequences: MATH 335 into MATH 370, MATH 120 into STAT 300, CHEM 300 into CHEM 400, and BIOL 430 into BIOL 431. In each case, the cumulative success rate of the instructor in the prerequisite course was used as a predictor of the student's likelihood of success in the next course. This analysis was conducted in order to uncover a potential correlation between an instructor's success rate in a prerequisite and later success for a student – with the primary goal of providing a framework for thinking about calibration and instructional improvement. Methods, analyses, and interpretation are further described in the next sections.

[3] Several lifetimes of research, probably.

Methods

Two primary analyses were conducted using student records. The first tested the association between a student's probability of success in a given course and the success rate in their prerequisite course. The second analysis rolled data up to the level of the *instructor* – correlating success rates for the instructor in prerequisite courses with success rates of their students in the next level. Taken together, these analyses are used to highlight a potential framework for thinking about calibration and instructional improvement.

Analysis 1 – Participants and Data Collection

As previously stated, the first analysis focused on projecting a student's probability of success. For the purposes of this investigation, four large STEM course sequences were identified – MATH 335 into MATH 370, MATH 120 into STAT 300, CHEM 300 into CHEM 400, and BIOL 430 into BIOL 431. Students who attempted the second course in each sequence for the first time from Fall 2012 to Fall 2022 were tracked backwards, and data were gathered on completion of their prerequisite course. Students who did not complete a prerequisite course at CRC were not included in the analysis. Moreover, students who took their sequential course with the same instructor as their prerequisite course were also not included in the analysis. The final student headcount and prerequisite instructor count can be found in *Table 1*. The MATH 120 into STAT 300 pathway was by far the largest, followed by MATH 335 into 370, CHEM 300 into CHEM 400, and BIOL 430 into BIOL 431. Note that the number of instructors in *Table 1* refers to the number that taught prerequisite courses. That is, the 2565 students included in the MATH 120 into STAT 300 sample had 75 different prerequisite instructors.

Table 1. Cohort Sizes

Course	N Students	N Prerequisite Instructors
MATH 335 -> MATH 370	978	37
MATH 120 -> STAT 300	2565	75
CHEM 300 -> CHEM 400	474	20
BIOL 430 -> BIOL 431	506	16

When conducting an analysis such as this, i.e. looking at the association between variables in a non-experimental setting, it is important to statistically control for other potential explanations. For example, it may be the case that an instructor with high success rates in the second level of a sequence passes more high-income students. Analyses should therefore hold constant, or "control", for variables that could act as an alternative explanation for the association of interest. That way, one can be more certain that the correlation/association of interest isn't explained by other intersecting factors.

To that end, demographic data were gathered for each student as potential control variables for the analysis. Specifically, data were gathered on race/ethnicity, gender, income, and age for each student during the term where they first attempted the second-sequential course. Students were classified as below-poverty, low-income, and middle income (or higher) on the basis of federal poverty standards. In addition to demographic variables, an academic success score (heretofore referred to as *success score*) was calculated as a control variable. The calculation of a success score proceeded like a GPA calculation. However, unlike GPA, transcript notations that are not normally included in the calculation (e.g. “P”, “W”, “EW”, “MW”, “IP”) were included. A “P” or “CR” notation was assigned two “grade” points, consistent with a “C” grade in GPA calculation. “W”, “EW”, “MW”, and “I” grades were assigned zero points. All other notations had the traditional number of grade points assigned (e.g. an “A” grade had 4 points assigned).

Table 2. Success score points by grade notation.

Notation	Points
A	4
B	3
C, P, CR	2
D	1
F, W, EW, I	0

Because the calculation of the success score proceeded like a GPA calculation, each student’s score was constrained between 0 and 4.0. This calculation used student data up-to and including the term where the prerequisite course was attempted - excluding data from the prerequisite course. This provided a measure of the academic success of a student independent of the sequence at the center of the analysis.

Two additional variables were gathered specifically related to the prerequisite course. First, the student’s grade from the prerequisite course was gathered (“A”, “B’, or “C”) as a control variable. Second, for each student, the prerequisite instructor’s cumulative success rate up-to and including the term of the prerequisite attempt was calculated. That is, if a student took MATH 120 in Fall 2019, a cumulative success rate for their instructor for MATH 120 was calculated up-to and including Fall 2019. This variable will be referred to as *cumulative instructor success* throughout the proceeding sections.

Headcounts and percentages for demographic groups by course can be found in *Table 3* below. *Table 4* contains means and standard deviations for the previously described continuous/numeric measures – age, success score, and cumulative instructor success.

Table 3. Summary of Categorical Demographic Predictors

	Math 370		STAT 300		CHEM 400		BIOL 431	
Demographic	N	%	N	%	N	%	N	%
African American	47	4.8%	181	7.1%	15	3.2%	30	5.9%
Asian	412	42.1%	840	32.7%	169	35.7%	163	32.2%
Filipino	68	7.0%	170	6.6%	30	6.3%	61	12.1%
Hispanic/Latino	188	19.2%	634	24.7%	127	26.8%	95	18.8%
Multi-Race	55	5.6%	159	6.2%	24	5.1%	23	4.5%
Native American	3	0.3%	5	0.2%	1	0.2%	0	0.0%
Other Non-White	4	0.4%	8	0.3%	3	0.6%	1	0.2%
Pacific Islander	9	0.9%	50	1.9%	6	1.3%	16	3.2%
Unknown	9	0.9%	17	0.7%	3	0.6%	6	1.2%
White	183	18.7%	501	19.5%	96	20.3%	111	21.9%
Female	303	31.0%	1470	57.3%	226	47.7%	362	71.5%
Male	661	67.6%	1062	41.4%	245	51.7%	135	26.7%
Unknown	14	1.4%	32	1.2%	3	0.6%	9	1.8%
Below Poverty	312	31.9%	886	34.5%	156	32.9%	149	29.4%
Low	256	26.2%	665	25.9%	112	23.6%	134	26.5%
Middle and Above	304	31.1%	722	28.1%	142	30.0%	179	35.4%
Unable to Determine	106	10.8%	292	11.4%	64	13.5%	44	8.7%
Total	978		2565		474		506	

Table 4. Summary of Continuous Predictors

	Math 370		STAT 300		CHEM 400		BIOL 431	
Demographic	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Success Score	2.86	0.83	2.74	0.77	2.73	0.84	2.89	0.78
Age	21.90	5.08	23.71	6.60	22.82	5.19	25.40	6.25
Cumulative Instructor Success	0.48	0.15	0.54	0.16	0.57	0.12	0.68	0.10

Analysis 2 – Participants and Data Collection

As previously stated, the second analysis correlated the success rate for an instructor in their prerequisite course with the success rate of their students in the next level. It used the same dataset as the first analysis but looked at the level of the instructor. For example, instead of looking at the 2565 students in the STAT pathway, the analysis focused on data for 75 instructors that had taught the prerequisite (MATH 120) for that sample of students. For each instructor, demographic/success data from the students who had passed their prerequisite course were aggregated. That is, for a given instructor, the analysis looked at the average success score of their students that passed the prerequisite with them, the percentage of low-income students amongst those that passed the prerequisite with them, etc. These were used as control variables in the analysis testing the association between success rate in the prerequisite and success rate in the second-sequential course. For this analysis, only MATH 120 into STAT 300 and MATH 335 into MATH 370 were included. The other sequences did not have enough instructors of the prerequisite to test statistical hypotheses at this level.

Because data for each instructor were rolled up into averages (e.g. average success score), it was important to ensure that each instructor had sufficient sample size of students in the second level. As such, instructors with fewer than five students attempting the sequential course for the first time within the ten-year time span were not included in this analysis. This helped winnow extreme values like 100% or 0% success rates in the second course. After filtering, a total of 33 instructors remained in the MATH 335 into MATH 370 sample and 60 remained in the MATH 120 into STAT 300 sample. A complete breakdown of the variables along with means (and standard deviations) can be found in *Table 5* below.

Table 5. Analysis 2 Means and Standard Deviations

	STAT 300		MATH 370	
Variable	Mean	SD	Mean	SD
Next Level Success Rate	0.59	0.14	0.51	0.18
Avg. Success Score	2.70	0.26	2.84	0.27
First Level Success Rate	0.53	0.17	0.47	0.18
% Low-Income	0.61	0.12	0.56	0.15
% African American	0.07	0.07	0.05	0.05
% White	0.19	0.09	0.20	0.13
% Hispanic/Latino(a)	0.24	0.09	0.20	0.09
% API	0.42	0.12	0.48	0.16
% Female	0.56	0.11	0.28	0.09

Analysis Procedure

The first analysis used a logistic regression with binomial error in order to test for a significant association between a student's probability of success in the next level and the cumulative success rate of the prerequisite instructor. Logistic regressions with binomial error are typically used in circumstances where there is an outcome variable with two discrete values. In this case, a student could either be successful or unsuccessful in the next level. The analysis proceeded in two steps. First, a logistic regression was conducted using the previously mentioned control variables. Second, the variable of interest – cumulative success rate – was entered into the model along with all control variables. This helped determine the association of cumulative success rate while holding all other variables constant (e.g. student income, student success score, etc.).

When testing for an association (say the association between cumulative instructor success and success in the next level), it is important to have an accurate estimate of how much that association may vary from sample-to-sample. Specifically, having an accurate and unbiased estimate of this *standard error* is essential for drawing conclusions about statistical significance. The *regression coefficient* – a description of how much the outcome variable changes on the basis of a one unit change in a predictor variable – is typically divided by this standard error to produce a test statistic[4]. Significance is determined on the basis of this test statistic. If, for some reason, the standard error is not estimated properly, then one might draw incorrect conclusions about the association in question.

It is therefore important to note that the data in the first analysis are inherently clustered. That is, students take classes in sections together, and as such they may be similar to each other in some way. In most circumstances, this clustering results in estimates of standard error that are too low, potentially resulting in spurious significant results. As such, cluster robust estimates of standard errors[5] were used in testing significance of associations. Errors were estimated assuming student data were clustered in their prerequisite course and their course at the next level.

The second analysis used ordinary least squares regression with success rate in the next level as the outcome variable. Demographic averages (*Table 5*) were entered as control variables in the first step, and in the second step the instructor's overall success rate in the prerequisite was entered. This helped determine significance of the association between success rate in the first level and success rate in the second level while holding other factors constant.

Prior to analysis, continuous variables in the first analysis (e.g. cumulative instructor success, success score, age) were mean centered and divided by their standard deviation. In the second analysis, all variables were mean centered and divided by their standard deviation. Regression coefficients should therefore be interpreted as the amount of change in the outcome variable *for one standard deviation of the continuous variable.*

Results

Analysis 1 – Association of Cumulative Instructor Success

Success rates in the subsequent course for each student demographic can be found in *Table 6* below. Continuous variables (age, cumulative instructor success, and success score) were divided into quartiles and success rate was calculated for each quartile band. These success rates can be found in *Table 7a* – Table *7d* for MATH 335 into MATH 370, MATH 120 into STAT 300, CHEM 300 into CHEM 400, and BIOL 430 into BIOL 431, respectively. Note that the sample

[4] This is intended to provide an intuition of the issue being described – e.g. statistical inference and standard error. This is not intended to be a textbook description of significance testing and nuances therein.

[5] e.g., C., A., Collin & D., L., Miller (2015). A Practitioner's Guide to Cluster-Robust Inference. *The Journal of Human Resources 50*(2). You may also find it useful to google the "sandwich estimator" – which sounds more delicious than it is.

size within quartile band varies due to repeat values at the lower quartile, median, and upper quartile cut-offs. The demographic variables listed in *Table 6* along with age, success score, and prerequisite grade were entered in the first step of the analysis.

Table 6. Subsequent Course Success Rate by Categorical Demographic Group

Variable	MATH 370	STAT 300	CHEM 400	BIOL 431
African American	38.3%	45.9%	33.3%	73.3%
Asian	57.0%	66.2%	75.7%	75.5%
Filipino	47.1%	65.3%	73.3%	77.0%
Hispanic/Latino	42.0%	54.4%	60.6%	82.1%
Multi-Race	43.6%	54.1%	62.5%	91.3%
Native American	66.7%	80.0%	0.0%	
Other Non-White	75.0%	75.0%	66.7%	100.0%
Pacific Islander	66.7%	60.0%	50.0%	68.8%
Unknown	88.9%	64.7%	66.7%	66.7%
White	53.6%	67.1%	65.6%	80.2%
Female	55.1%	61.7%	64.2%	76.8%
Male	50.1%	60.7%	69.4%	83.0%
Unknown	50.0%	46.9%	66.7%	66.7%
Below Poverty	51.0%	59.5%	64.1%	75.2%
Low	56.6%	59.8%	66.1%	77.6%
Middle And Above	49.7%	65.0%	74.6%	82.7%
Unable to Determine	47.2%	59.6%	57.8%	72.7%
Total	**51.6%**	**61.1%**	**66.9%**	**80.4%**

Table 7a. Subsequent Course Success Rates for Continuous Variables - MATH 370

Variable	Range	Success Rate	N
Age	<= 19	50.4%	343
	19.01-20	50.8%	193
	20.01-23	47.7%	243
	23 <	59.3%	199
Success Score	<= 2.3	29.8%	245
	2.31-2.96	41.4%	244
	2.97-3.52	59.4%	244
	3.52 <	75.9%	245
Cumulative Instru	<= 37.5%	61.4%	251
	37.6% -45.5%	58.0%	238

	45.6%-54.3%	49.4%	247
	54.3% <	37.6%	242

Table 7b. Subsequent Course Success Rates for Continuous Variables - STAT 300

Variable	Range	Success Rate	N
Age	<= 20	62.3%	962
	20.01-21.00	58.2%	383
	21.01-25.00	57.2%	645
	25 <	65.6%	575
Success Score	<= 2.19	46.4%	642
	2.20-2.79	52.0%	641
	2.80-3.33	62.8%	659
	3.33 <	83.9%	623
Cumulative Instructor Success	<= 40.5%	70.4%	642
	40.5%-55.8%	64.0%	648
	55.8%-66.3%	55.8%	661
	66.3% <	54.1%	614

Table 7c. Subsequent Course Success Rates for Continuous Variables - CHEM 400

Variable	Range	Success Rate	N
Age	<= 20	71.1%	197
	20.01-21	67.8%	90
	21.01-24	60.5%	76
	24 <	63.1%	111
Success Score	0-2.16	44.2%	120
	2.17-2.77	57.3%	117
	2.78-3.39	75.4%	118
	3.39-4	90.8%	119
Cumulative Instructor Success	<= 46.8%	69.0%	126
	46.9%-59.6%	77.3%	119
	59.7%-63.6%	65.3%	118
	63.6% <	55.0%	111

Table 7d. Subsequent Course Success Rates for Continuous Variables - BIOL 431

Variable	Range	Success Rate	N
Age	<= 21	80.6%	160
	21.01-23	76.8%	112
	23.01-28	80.2%	116
	28 <	74.6%	118
Success Score	<= 2.36	63.0%	127
	2.37-2.95	70.6%	126
	2.96-3.48	85.8%	127
	3.48 <	93.7%	126
Cumulative Instructor Success	<= 62.1%	87.1%	132
	62.2%-66.4%	77.2%	127
	66.5%-72.2%	76.7%	120
	72.2% <	71.7%	127

After controlling for race, income, gender, age, grade in the prerequisite, and success score, instructor cumulative success rate was a significant predictor of success in the sequential course. On average, higher success rates in the prerequisite course predict a lower probability of success in the next level, $z = -7.64$, $p < .001$, $z = -6.83$, $p < .001$, $z = -4.12$, $p < .001$, $z = -3.72$, for MATH 335 into MATH 370, MATH 120 into STAT 300, CHEM 300 into CHEM 400, and BIOL 430 into BIOL 431, respectively. Parameter estimates and significance tests can be found in *Table 8* below. Note that slope estimates for binomial logistic regression can interpreted as the change in the log of the odds of success for a one unit change in the independent variable (which is measured in standard deviations). The "Adjusted *p*-value" column indicates the significance of the association after adjusting standard errors for clustering.

For the purposes of interpretation, the odds ratio provides a measure of the relative "size" of the association. In this case, odds ratios were calculated on the basis of a one-standard deviation change in cumulative instructor success from the average cumulative instructor success. In this case, a smaller value of the odds ratio indicates a larger decline in the probability of success. MATH 120 into STAT 300 had the weakest association whereas MATH 335 into MATH 370 exhibited the strongest. Full model parameter estimates, odds ratios, and standard errors for each sequence can be found in the appendix of this report.

Table 8. Regression Slopes and Odds Ratios for The Effect of Cumulative Success Rate

Sequence	Slope Estimate	Odds Ratio	Standard Error	*z*	*p*-value	Adjusted *p*-value
MATH 335 into 370	-0.66	0.52	0.09	-7.64	< .001	< .001

MATH 120 into STAT 300	-0.32	0.73	0.05	-6.83	< .001	< .001
CHEM 300 into CHEM 400	-0.46	0.63	0.11	-4.12	< .001	< .001
BIOL 430 into BIOL 431	-0.41	0.66	0.11	-3.72	< .001	< .001

Analysis 1 – Model Accuracy and Prediction Error

The odds ratios of an association give a good measure of the size of an "effect" or association. Another way to look at the effectiveness of a statistical model – like the logistic regressions described in the last section – is to look at how accurately it predicts the outcome variable. The data for a student (age, success score, prerequisite grade, etc.) can be entered into the logistic regression model in order to project their probability of success in the next level. A projected probability above 50% for a student indicates that, on average, they are more likely to succeed than not. In the case of a projected probability of 50% or more, the statistical model would therefore predict successful completion of the next level.

Projections for each student can then be compared to the actual success of the student in the next level for a measure of accuracy. The percentage of correct guesses by each logistic regression (four in total) can be found in *Table 9* below. It is particularly useful to compare these model accuracy estimates to the overall success rate in the sequential course. For example, the success rate in STAT 300 for the whole sample of students was 61.1%. With no other information, this means that a student (from this sample) had a 61.1% chance of succeeding in the course. A model might therefore guess that every student is likely to succeed – a guess that would be correct 61.1% of the time (the success rate of all students). With the additional information added by the logistic regression model for MATH 120 into STAT 300, the accuracy improves to 68%. The largest relative increase was observed for MATH 335 into MATH 370 – an improvement of about 20% (72% from the logistic regression model - 51.6% from the average success rate).

Table 9. Model Prediction Accuracy

Model	Prediction Accuracy
MATH 335 into MATH 370	72.0%
MATH 120 into STAT 300	68.0%
CHEM 300 into CHEM 400	75.9%
BIOL 430 into BIOL 431	80.4%

There are indeed many ways to evaluate the predictive capacity of a model. For example, one might look at the model precision (e.g. the percentage of successful students who were correctly identified out of the total number who were projected to succeed)[6]. This discussion of model accuracy is, however, important simply to assert that the statistical models in the first analysis are imperfect (as is the case with the vast majority of statistical models). As the second analysis will illustrate, the errors in prediction may help develop an *initial* model for considering calibration and instructional improvement.

Analysis 2 – Association of Success Rate in the First Level with Success Rate in the Second Level

As previously stated, the second analysis tested the association between a prerequisite instructor's success in the first level and the success rate of their students in the second level. In the first step of this analysis, control variables were entered (e.g. average success score, percent low-income, etc.) and in the second step, the instructor's success rate for the prerequisite course was entered as a predictor. In both sequences, success rate for the prerequisite instructor was a significant predictor of the success rate in the second-sequential course. The association was stronger for the MATH 335 into MATH 370 sequence, $t(1) = -4.31$, $p < .001$, $R^2_{partial} = 0.45$, than MATH 120 into STAT 300, $t(1) = -2.62$, $p < 0.05$, $R^2_{partial} = 0.12$. On the basis of partial R^2 – a measure of effect size – the effect in the MATH 370 pathway would be considered "large" and the effect in the STAT 300 pathway would be considered "medium".

Table 10. Regression Slopes and Significance - MATH 335 into MATH 370

Predictor	Slope Estimate	Standard Error	*t*	*p*-value
Intercept	0.51	0.02	23.50	< .001
% Female	-0.01	0.02	-0.51	*ns.*
Average Age	0.04	0.02	1.53	*ns.*
% API	-0.02	0.09	-0.22	*ns.*
% Hispanic/Latino	-0.08	0.05	-1.38	*ns.*
% White	0.01	0.07	0.17	*ns.*
% African American	0.02	0.03	0.54	*ns.*
% Low Income	0.03	0.03	0.88	*ns.*
Average Ability Score	0.03	0.03	0.85	*ns.*
Average Success in MATH 335	-0.10	0.02	-4.31	< .001

Table 11. Regression Slopes and Significance - MATH 120 into STAT 300

Predictor	Slope Estimate	Standard Error	*t*	*p*-value
Intercept	0.59	0.02	36.58	< .001
% Female	0.01	0.02	0.36	*ns.*
Average Age	-0.01	0.02	-0.53	*ns.*
% API	0.06	0.03	1.74	*ns.*
% Hispanic/Latino	0.02	0.03	0.71	*ns.*
% White	0.03	0.03	1.05	*ns.*

[6] A logistic regression would have substantially better precision than a projection based exclusively on the success rate alone.

% African American	0.02	0.02	0.72	*ns.*
% Low Income	-0.03	0.02	-1.76	*ns.*
Average Ability Score	-0.01	0.02	-0.42	*ns.*
Average Success in MATH 120	-0.06	0.02	-2.62	< .05

The finding from the second analysis mirrors the findings from the first analysis. There is a negative average association between an instructor's success rate in the first course of a sequence and the success rate of their students in the second course.

Analysis 2 – Model Accuracy and Prediction Error

Like the models described for the first analysis of this report, it is important to point out the imperfectness of the association between an instructor's success in the first level and the success of their students in the next. This association explained about 45% of the remaining variance in the MATH 335 pathway after controlling for other variables and about 12% in the MATH 120 pathway. As such, both associations are clearly visible on a plot. Nevertheless, some instructors have relatively low success rates in the first course and relatively low success rates in the next level. Conversely, some instructors have relatively high success rates in the first course and relatively high success rates in the next level. There is error in prediction because some data are not perfectly described by the aforementioned negative association.

To illustrate this concept, the association between instructor success in the first level and the success of their students in the next level has been plotted in *Figure 1* and *Figure 2* below. The horizontal and vertical lines on the plot represent the median success rate in the second level and the median success rate in the first level, respectively. The dotted line represents the projected success rate in the second level at the average values of the control variables for each regression analysis. The size of each dot represents the sample size for an instructor (the number of students who attempted the next level between Fall 2012 and Fall 2022) and the color represents the average success score of those students.

Figure 1. MATH 335 into MATH 370

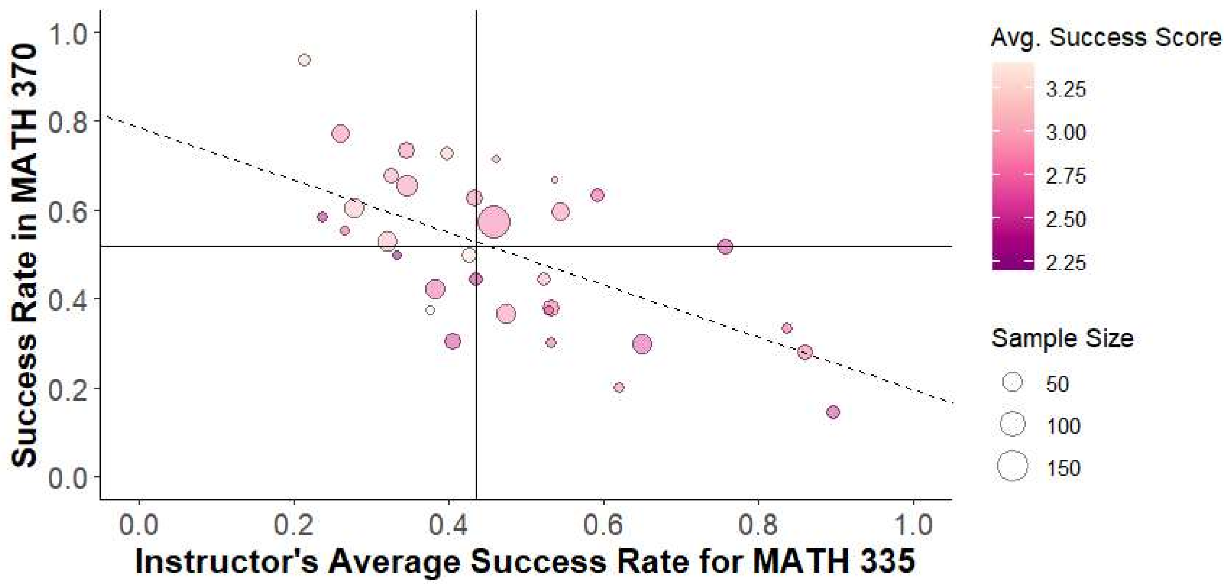
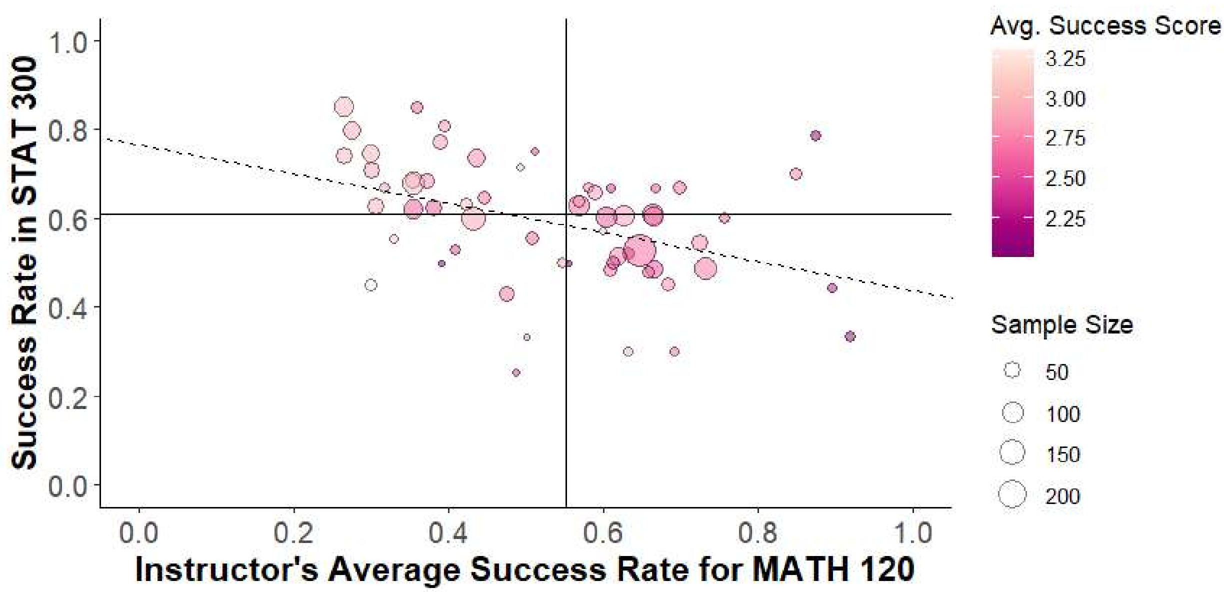


Figure 2. MATH 120 into STAT 300



Inspecting these plots is a useful exercise in understanding how data points (representing instructors) may not be perfectly fit by the average downward trend. For example, in both plots there are data points in the upper right-hand quadrant. These are instructors who have relatively high success rates in the prerequisite (above the median) but still have a higher than median success rate in the next level. A similar observation can be made for instructors in the lower left-hand quadrant.

Another way to look at this variability is by examining differences in success rate for the second course while holding success rate in the first course constant. *Figure 3* and *Figure 4* below highlight a set of data points within a small range of success in a prerequisite course. The highlighted points can be found between thick vertical lines on each plot. Despite having similar success rates in the prerequisite course, some points are higher than others even though they span a small range of course success. Indeed, the purpose of highlighting the error in prediction here is to make an important point about success rates in sequential courses. The error in prediction – although in the context of strong/medium effect sizes – does not make it a foregone conclusion that an instructor with a relatively high success rate will have a low success rate in the next level, and vice versa. Prior to drawing conclusions about the meaning of the success rate in the first level, the second success rate should also be considered.

Figure 3. MATH 335 into MATH 370

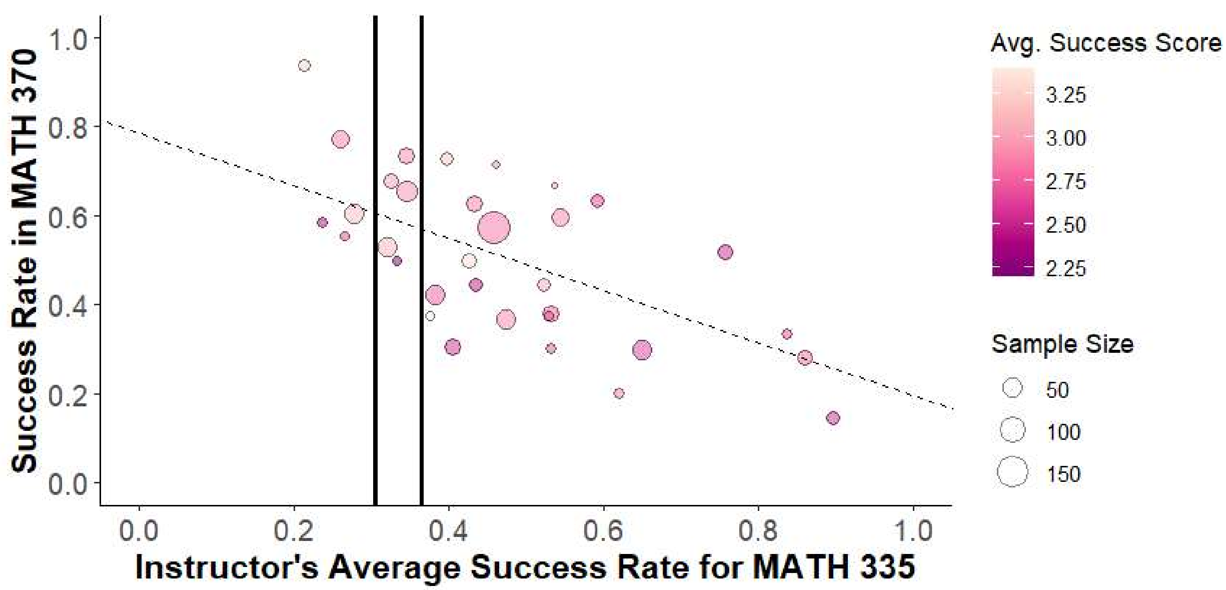
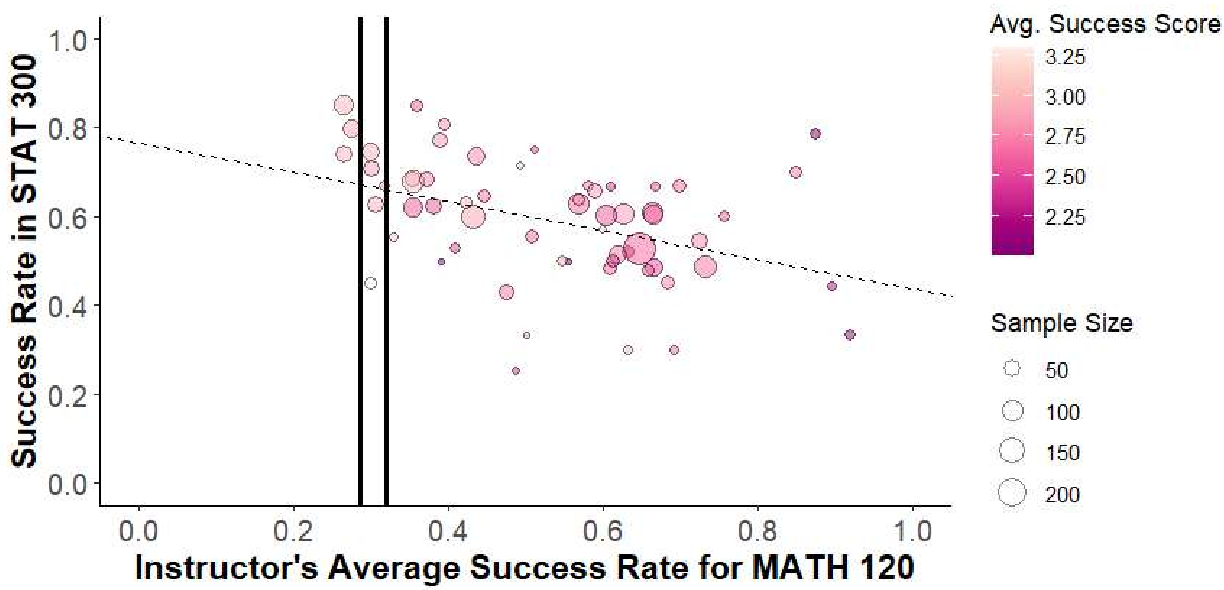


Figure 4. MATH 120 into STAT 300



Discussion and Conclusions

The present investigation found a negative correlation between the success rate of a prerequisite instructor and the success of their students in the next level. However, the size of the effect varied depending on the STEM sequence being investigated. MATH 120 into STAT 300 exhibited the weakest association, whereas MATH 335 into MATH 370 exhibited the strongest.

More importantly, the examination of this association and errors in prediction therein may suggest an initial (perhaps imperfect) framework for thinking about calibration and instructional improvement in STEM sequences.

Calibration

When thinking about an initial definition of calibration, it is important to consider the negative correlation uncovered by analysis 1 and analysis 2. Specifically, on average, there seems to be a trade-off between course success in the prerequisite and course success in the sequential course. That said, an instructor may be more certain that they have "calibrated" if there has been a similar trade-off between success in their first course and the success of their students in the second. For example, suppose an instructor increases the success rate of their STEM course and finds that students don't do quite as well in the next level. This may be seen as a sign of calibration. That is to say, one success rate increased (in this case the prerequisite course success rate) and the other success rate decreased (in this case the success rate in the sequential course).

Instructional Improvement

As previously stated, calibration considered the negative correlation between success rates in sequential courses. When thinking about instructional improvement, on the other hand, it may be important to consider the *error* in prediction discussed in analysis 1 and analysis 2. Specifically, there exist instructors who are above the median success rate in the first course of the sequence, and their students have an above median success rate in the next level. Because their data exists in the upper right quadrant of the plot (e.g. *Figure 1* and *Figure 2*), they are not perfectly described by the negative association in both analyses. Indeed, it is useful to apply this observation to the conceptualization of instructional improvement. That is to say, instructional improvement would *also* move counter to the negative association described in analysis 1 and 2. For example, suppose an instructor increases their course success rate (e.g. by some intervention), *and* they observe an increase in the success of their students in the next level. They may be more confident that they have improved their instruction because – not only have they increased their course success – their students are achieving a higher standard in the next level. Similarly, suppose an instructor increases their course success in the first level, and their students do equivalently in the next (e.g. no increase in course success in the next level). They may be more confident that they have improved their instruction because they have successfully moved more students to their standards. Additionally, and similarly, one might draw similar conclusions if an instructor increased success in the next level without changing their course success in the first level. In all three of the aforementioned scenarios, change in course success is not accompanied by a trade-off, running counter to calibration. It may therefore be useful to initially define instructional improvement as:

1. An increase in success in the prerequisite course accompanied by equivalent and/or increased success in the second-level sequential course.
2. Or, an increase in success in the second-sequential course accompanied by steady success in the prerequisite course.

An example of this conceptualization can be found in *Figure 5* below. The point in the middle of the plot is the initial data for a course or instructor (prior to some intervention) - 50% success in the first course and 50% success in the subsequent course. Two lines are drawn through the point in the center to break the plot up into quadrants. After some instructional change, the instructor observes that they have increased their course success to 60% and now 60% of their passing students are successful in the next level. That is, their data has moved into the upper right-hand quadrant (as indicated by the arrow in the plot). The instructor may be more certain they have improved learning for their students. Not only have they increased their course success – their students are achieving a higher standard in the next level.

Figure 5. Example Instructional Improvement.

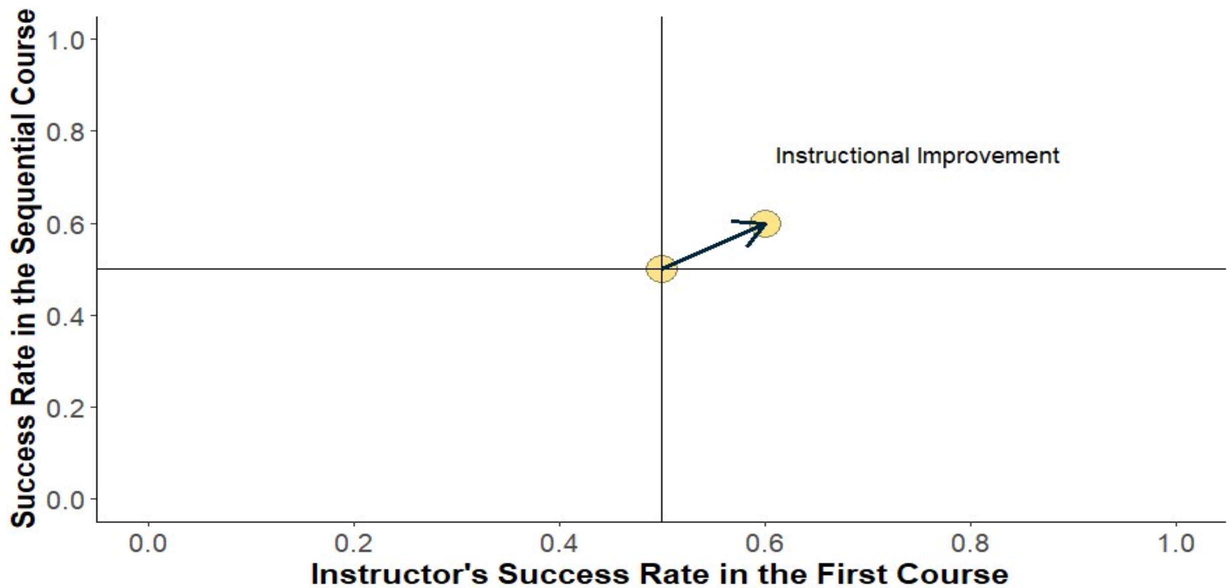
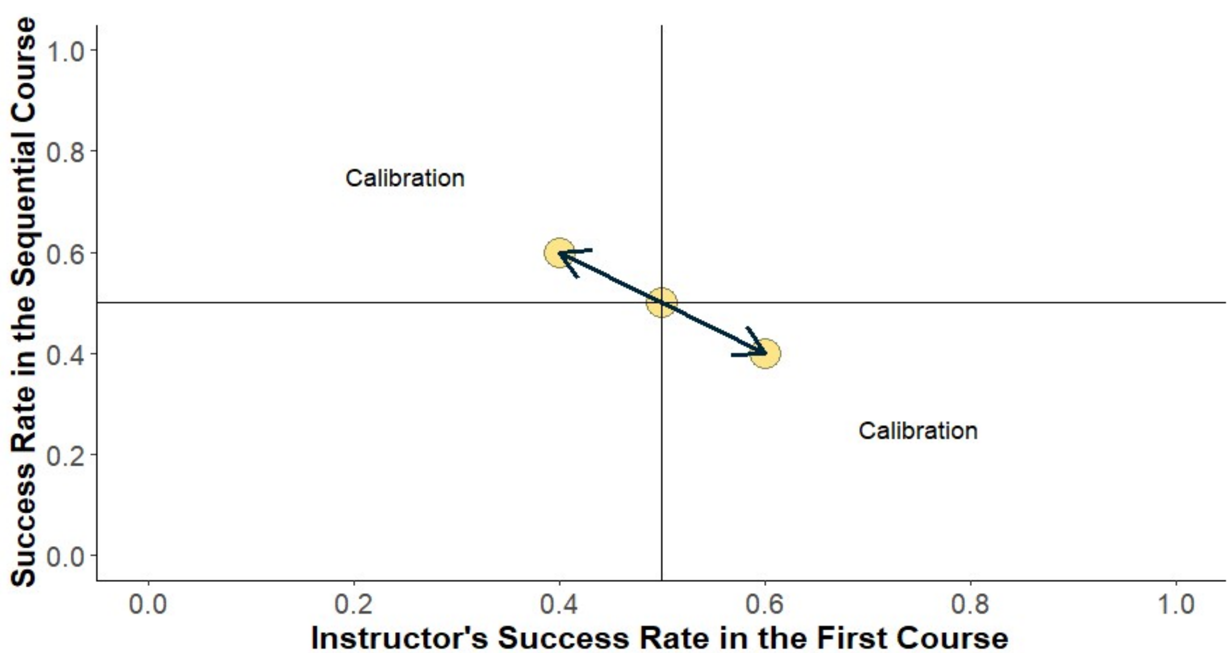


Figure 6. Example of Calibration.



Considerations of the Working Definitions of Instructional Improvement and Calibration

It is important to emphasize that the definitions proposed here are imperfect and should be treated as an *initial* framework for thinking about calibration and instructional improvement. For example, it is not hard to see how an increase in success in the first and second level might reflect some combination of instructional improvement and calibration (as defined in the first section of this report). That is, the instructional change may have resulted in increased learning and changes to standards. However, the relative impact of both averaged out to movement into the upper-right quadrant. It is also possible for the opposite to happen, e.g. a change in instruction and standards may result in movement in success rates that is consistent with operational definition of calibration proposed here.

Other factors not immediately related to the instructor may also induce movement of success rates. For example, changes in the standards of instructors in the next level may result in success rates changing. For an instructor reflecting on their data, this would have very little meaning for instructional improvement or calibration. Nevertheless, an instructor might find more comfort in using the framework here if they are less comfortable working with success rates alone.

Finally, the findings reported here only apply to sequential courses. It is difficult to tell how one might distinguish between instructional improvement and calibration in a single course. This topic is unfortunately beyond the scope of the study presented here.

Conclusions and Recommendations

The two analyses of this study found negative correlations between success in the first course and success in the second course of a STEM sequence. Consideration of these associations and errors therein led to two working definitions of instructional improvement and calibration – the primary goal of this investigation. On the basis of the findings reported here the Research Office at CRC makes the following recommendations:

1. CRC instructors within course sequences should consider using *both* success in the prerequisite course and success in the sequential course to assess instructional improvement or calibration.
2. Efforts to increase course success should emphasize and provide resources towards instructional improvement.
3. The Research Office at CRC should provide data visualization for sequential course success data to support instructors within sequences to discern improvement in their instruction.
4. Future research should also focus on the joint probability of a student passing the first and second course in a sequence and the factors (instructional or otherwise) that maximize that probability.

Appendix

Note: Age, ability score, and cumulative instructor success were each standardized by their mean and standard deviation prior to analysis.

MATH 335 into MATH 370 Regression Analysis Parameters and Significance Tests

Predictor	Slope Estimate	Odds Ratio	Standard Error	*z*	*p*-value	Adjusted *p*-value
Intercept	1.22	3.38	0.41	2.95	< .01	< .05
Age	0.12	1.13	0.10	1.26	ns.	ns.
Race/Ethnicity (African American Baseline)						
Asian	0.29	1.34	0.36	0.82	ns.	ns.
Filipino	0.00	1.00	0.43	0.00	ns.	ns.
Hispanic/Latino	0.01	1.01	0.38	0.02	ns.	ns.
Multi-Race	0.12	1.13	0.45	0.27	ns.	ns.
Native American	1.68	5.36	1.36	1.24	ns.	ns.
Other Non-White	1.40	4.06	1.28	1.09	ns.	ns.
Pacific Islander	1.40	4.07	0.96	1.46	ns.	ns.
Unknown	2.84	17.11	1.41	2.01	< .05	< .01
White	0.35	1.42	0.38	0.93	ns.	ns.
Gender (Female Baseline)						
Male	0.12	1.13	0.17	0.71	ns.	ns.
Unknown	0.13	1.14	0.65	0.20	ns.	ns.
Success Score	0.48	1.61	0.09	5.60	< .001	< .001
Prior Course Grade (A Grade Baseline)						
B	-1.52	0.22	0.25	-6.19	< .001	< .001
C	-2.34	0.10	0.25	-9.25	< .001	< .001
Income Level (Poverty-Level Baseline)						
Low-Income	0.44	1.55	0.20	2.16	< .05	< .05
Middle Income and Above	0.10	1.11	0.20	0.53	ns.	ns.
Unable to Determine	-0.04	0.96	0.27	-0.16	ns.	ns.
Cumulative Instructor Success	-0.66	0.52	0.09	-7.64	< .001	< .001

MATH 120 into STAT 300 Regression Analysis Parameters and Significance Tests

Predictor	Slope Estimate	Odds Ratio	Standard Error	*z*	*p*-value	Adjusted *p*-value
Intercept	0.60	1.82	0.20	2.95	< .01	< .01
Age	0.04	1.04	0.05	0.82	ns.	ns.
Race/Ethnicity (African American Baseline)						
Asian	0.59	1.81	0.18	3.30	< .001	< .01

Filipino	0.60	1.83	0.24	2.56	< .05	< .01
Hispanic/Latino	0.28	1.32	0.18	1.53	ns.	ns.
Multi-Race	0.10	1.11	0.24	0.43	ns.	ns.
Native American	1.20	3.31	1.19	1.00	ns.	ns.
Other Non-White	1.24	3.47	0.87	1.42	ns.	ns.
Pacific Islander	0.58	1.78	0.35	1.65	ns.	ns.
Unknown	0.58	1.79	0.57	1.02	ns.	ns.
White	0.56	1.75	0.19	2.97	< .01	< .01
Gender (Female Baseline)						
Male	0.10	1.11	0.09	1.12	ns.	ns.
Unknown	-0.66	0.52	0.39	-1.69	ns.	ns.
Success Score	0.50	1.64	0.05	10.12	< .001	< .001
Prior Course Grade (A Grade Baseline)						
B	-0.58	0.56	0.13	-4.52	< .001	< .001
C	-1.06	0.35	0.13	-8.13	< .001	< .001
Income Level (Poverty-Level Baseline)						
Low-Income	0.08	1.08	0.11	0.72	ns.	ns.
Middle Income and Above	0.23	1.26	0.12	2.02	< .05	ns.
Unable to Determine	0.08	1.08	0.15	0.54	ns.	ns.
Cumulative Instructor Success	-0.32	0.73	0.05	-6.83	< .001	< .001

CHEM 300 into CHEM 400 Regression Analysis Parameters and Significance Tests

Predictor	Slope Estimate	Odds Ratio	Standard Error	*z*	*p*-value	Adjusted *p*-value
Intercept	-0.15	0.86	0.71	-0.21	ns.	ns.
Age	-0.15	0.86	0.12	-1.22	ns.	ns.
Race/Ethnicity (African American Baseline)						
Asian	2.17	8.75	0.70	3.08	< .01	< .001
Filipino	1.72	5.57	0.82	2.10	< .05	< .01
Hispanic/Latino	1.45	4.26	0.69	2.09	< .05	< .001
Multi-Race	1.51	4.51	0.84	1.79	ns.	< .01
Other Non-White	1.07	2.92	1.52	0.71	ns.	ns.
Pacific Islander	0.95	2.60	1.07	0.89	ns.	ns.
Unknown	1.11	3.02	1.59	0.70	ns.	ns.
White	1.60	4.93	0.71	2.24	< .05	< .001
Gender (Female Baseline)						
Male	0.06	1.07	0.24	0.27	ns.	ns.
Unknown	-0.69	0.50	1.49	-0.47	ns.	ns.
Success Score	0.58	1.78	0.13	4.38	< .001	< .001
Prior Course Grade (A Grade Baseline)						
B	-0.91	0.40	0.31	-2.94	< .01	< .001

C	-1.90	0.15	0.34	-5.67	< .001	< .001
Income Level (Poverty-Level Baseline)						
Low-Income	0.08	1.08	0.11	0.72	ns.	ns.
Middle Income and Above	0.23	1.26	0.12	2.02	< .05	ns.
Unable to Determine	0.08	1.08	0.15	0.54	ns.	ns.
Cumulative Instructor Success	-0.46	0.63	0.11	-4.12	< .001	< .001

BIOL 430 into BIOL 431 Regression Analysis Parameters and Significance Tests

Predictor	Slope Estimate	Odds Ratio	Standard Error	*z*	*p*-value	Adjusted *p*-value
Intercept	2.87	17.59	0.64	4.49	< .001	< .001
Age	-0.13	0.88	0.15	-0.88	ns.	ns.
Race/Ethnicity (African American Baseline)						
Asian	-0.48	0.62	0.55	-0.87	ns.	ns.
Filipino	-0.41	0.66	0.60	-0.68	ns.	ns.
Hispanic/Latino	0.50	1.65	0.58	0.87	ns.	ns.
Multi-Race	1.47	4.35	0.96	1.52	ns.	ns.
Pacific Islander	-0.18	0.84	0.79	-0.22	ns.	ns.
Unknown	-0.53	0.59	1.18	-0.45	ns.	ns.
White	-0.11	0.90	0.56	-0.19	ns.	ns.
Gender (Female Baseline)						
Male	0.46	1.58	0.31	1.49	ns.	ns.
Unknown	-1.18	0.31	0.84	-1.40	ns.	ns.
Success Score	0.48	1.62	0.13	3.69	< .001	< .001
Prior Course Grade (A Grade Baseline)						
B	-1.11	0.33	0.38	-2.89	< .01	< .01
C	-2.40	0.09	0.40	-6.08	< .001	< .001
Income Level (Poverty-Level Baseline)						
Low-Income	-0.09	0.92	0.33	-0.27	ns.	ns.
Middle Income and Above	-0.06	0.95	0.33	-0.17	ns.	ns.
Unable to Determine	-0.61	0.54	0.46	-1.31	ns.	ns.
Cumulative Instructor Success	-0.41	0.66	0.11	-3.72	< .001	< .001