Boise State University
ScholarWorks

Mathematics Faculty Publications and Presentations

Department of Mathematics

1-1-2016

Longitudinal Success of Calculus I Reform

Doug Bullock
Boise State University

Kathrine E. Johnson

Janet Callahan
Boise State University

New Orleans Jazzed about Engineering Education

Longitudinal Success of Calculus I Reform

Dr. Doug Bullock, Boise State University

Doug Bullock is an Associate Professor of Mathematics at Boise State University. His educational research interests include impacts of pedagogy on STEM student success and retention.

Dr. Kathrine E. Johnson
Dr. Janet Callahan, Boise State University

Janet Callahan is Chair of Materials Science and Engineering at Boise State University. Dr. Callahan received her Ph.D. in Materials Science, M.S. in Metallurgy, and B.S. in Chemical Engineering from the University of Connecticut. Her educational research interests include freshman engineering programs, math education, K-12 STEM curriculum and accreditation, and retention and recruitment of STEM majors.

Longitudinal Success of Calculus I Reform

Abstract

This paper describes the second year of an ongoing project to transform calculus instruction at Boise State University. Over the past several years, Calculus I has undergone a complete overhaul that has involved a movement from a collection of independent, uncoordinated, personalized, lecture-based sections, into a single coherent multi-section course with an active-learning pedagogical approach. The overhaul also significantly impacted the course content and learning objectives. The project is now in its fifth semester and has reached a steady state where the reformed practices are normative within the subset of instructors who might be called upon to teach Calculus I. Gains from the project include a rise in the pass rate in Calculus I, greater student engagement, greater instructor satisfaction, a general shift toward active learning pedagogies, and the emergence of a strong collaborative teaching community.

Project leaders are seeking to expand these gains to other areas of the curriculum and to broaden the community of instructors who are fully accepting of the reforms. Common concerns expressed by faculty resistant to the overhaul include suspicion that pass rate gains might reflect grade inflation or weakened standards, and that altering the traditional content of Calculus I might leave students unprepared for Calculus II. External stakeholders also have a vested interest in ensuring students receive a solid preparation in Calculus I. In this paper we develop a response to ensure solid evidence of Calculus II readiness that we hope will be useful to change agents and campus leaders in many other settings.

We address concerns about Calculus II readiness by conducting a natural experiment, tracking two cohorts of students through Calculus I and into Calculus II. The "treatment" cohort consists of students who reach Calculus II after passing the reformed Calculus I. The "control" cohort consists of students who reach Calculus II after passing non-reformed Calculus I at Boise State University. The experiment has no designed randomizing, but enrollment data shows that both cohorts spread out across all sections of Calculus II with apparent randomness. Our research question is: "Does the treatment cohort perform any worse than the control cohort in Calculus II?" Data on pass rates and grades in Calculus II will show that the answer is "No."

Introduction: History of the Calculus I reform project.

Boise State University has been experiencing growth in STEM enrollment every year since the formation of its college of engineering in 1997. In fall, 2015, STEM enrollment included nearly 4,000 students. Accompanying this growth came a demand for increased capacity in Calculus I, which has grown in enrollment by 74% over the past decade (from 244 in fall, 2006 to an enrollment of 433 in fall, 2015).

With the increased demand for calculus instruction came several undesirable consequences. These included a lack of coherence between instructors in terms of content. Related to this was a lack of agreement in terms of what exactly students were expected to be able to do by the end of the course. In fact, that topic – the learning outcomes of the course – had not been addressed; each instructor instead carried their own learning outcomes. In nearly all instances, these outcomes were not actually articulated into a statement such as, "By the end of this course (chapter, section, unit), students will be able to…," but rather were internalized; each instructor had their own sense of what should be taught in calculus, which guided their teaching, assignments and examinations.

There was agreement about what c text should be used, and a common syllabus was on file. Yet, as a result of both growth and lack of coordination between instructional faculty, a situation had developed by 2005-6 which students, the mathematics department, and others recognized as being problematic. At that time, from a student's perspective, it appeared to matter more, "who you took," than "what you learned" in terms of their chances of passing the course.[1] This was supported by pass rate data; the average pass rate in 2005-6 was 51% and ranged from 30% to 90% depending on who taught the course.[2] The variation in pass rate was a confounding problem in post-requisite courses such as Calculus II; students had highly variable preparation, and the Calculus II course also had no framework of common learning outcomes.

In part as a result of an externally funded Science Talent Expansion Program (STEP) grant from the National Science Foundation in which the Chair of Mathematics was a co-investigator with the Director for the Center for Teaching and Learning, but also motivated internally by the mathematics department, and by the Office of the Provost, an initiative was launched to tackle calculus. This effort has been described elsewhere[1] and is briefly summarized below.

Our efforts at reform were influenced by a successful first-year engineering program at Wright State University[3] which focused on engineering applications of mathematics and also informed by faculty development research summarized by Bressoud, et. al.[4] In our reform, we were able to use STEP grant funding for year-long STEM-focused faculty learning communities (FLCs).[5] We held three FLCs across the five-year grant, with the last two cohorts exclusively focused on calculus instruction. These FLCs were facilitated by one calculus instructor who had reframed his calculus content into an application-based focus oriented to help future engineers and scientists appreciate why they need to learn calculus. A brief overview on the course is given below; full details on the FLCs, and how our "coherent calculus" course was developed, supported and implemented are presented elsewhere.[1]

Coherent Calculus -- overview

The "Coherent Calculus" course contains the following elements, outcomes and pedagogical approach:[1]

- Whenever possible, students work with data sets and/or continuous models selected from actual physical, biological, financial or other applied models.
- Whenever possible, Calculus concepts are introduced and motivated by application to these models and data sets.
- Whenever possible, content is presented using notation, language and conventions of the disciplines from which the models are taken.
- As much as possible, content will be relevant, recognizable, and applicable in subsequent STEM coursework.
- All content will be accessible from an intuitive or practical viewpoint. In particular, the level of abstraction will be significantly less than typically found in Calculus I.

Thematically the revised Calculus I class is focused on three outcomes:

- Develop geometric and physical intuition for derivatives and integrals.
- Master the standard rules for symbolic computation of derivatives and some basic integrals.
- Apply both intuitive understanding and rules mastery to solve problems.

The course design has the following pedagogical features:

- Many short homework assignments with immediate computer driven feedback/assessment, typically due on a two-day cycle.
- Each assignment designed along learning cycle principles to target one or two specific learning goals.
- The vast majority of class time devoted to students working in small groups on these homework assignments.
- Additional active learning assignments that occur in-class with real-time formative assessment (these were added in 2015-16).
- All in-class work facilitated by lead instructor and peer learning assistant
- Additional and more involved weekly work with written feedback.

Table 1 summarizes the successful adoption strategy. Long-lasting change for us was derived from an approach driven by the faculty, based on homework. That is, our reform was driven by what the calculus instructors agreed that students needed to be able to do, not from any sort of imposed model or framework.

Table 1: How to create a coordinated course with a common pedagogical approach

Adoption Strategy – NO – Do not do this!	Adoption Strategy – YES – Do this:
Appoint a course coordinator.	Start with common homework. Build consensus agreement on every exercise.
Combine 12 small sections into 2 huge sections.	Do the same thing with quizzes.
Impose a common syllabus.	Agree on basic weighting of all Hw, Qz, Ex, Final and letter-grade cutoffs.
Impose a common final.	Build consensus on exam content – eventually reach common exams.
Impose common midterm exams and final	Eventually adopt a common final
Impose a pedagogical model.	Along the way, allow the course content to shape pedagogy with the strategic goal, for example, of active learning.

The Impact of Coherent Calculus

Enrollment and Pass Rates: We now have two full years of pass rate data since the Calculus I project scaled up, see Figure 1. There was an initial jump from approximately 65% to 75% that has not been sustained, but the overall pattern is still good. Prior to scale up the historical pass rate was about 60% -- improving, but struggling to ever exceed 65%. Since the scale up pass rate has averaged 72% and only once dipped below 70%. Although participation in the project remains purely voluntary, all instructors continue to opt into the project except for those teaching honors or online sections.[i]
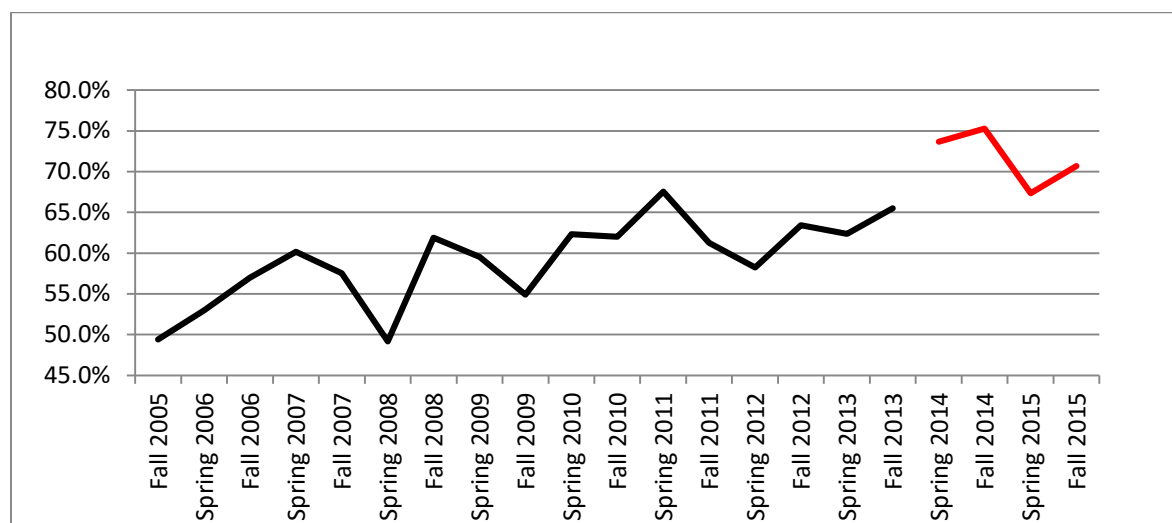

Figure 1: Calculus I pass rate as a function of semester. Red line corresponds to pass rates after the Coherent Calculus model was implemented in spring, 2014.

---

[i] In summer 2014 it was not possible to join, since materials were not ready. One summer 2015 instructor declined to participate. Pass rate and enrollment data here do not include summers. Summer does not much alter the overall averages or trends, but it is much more volatile.

The total number of students served by the reformed Calculus, compared to more traditional Calculus, is shown in Figure 2. During the initial scale up term, spring 2014, there were roughly equal numbers of reform and other Calculus sections. Presently, the only non-reform sections are (1) honors, (2) online, or (3) face-to-face but taught in parallel with the online section.
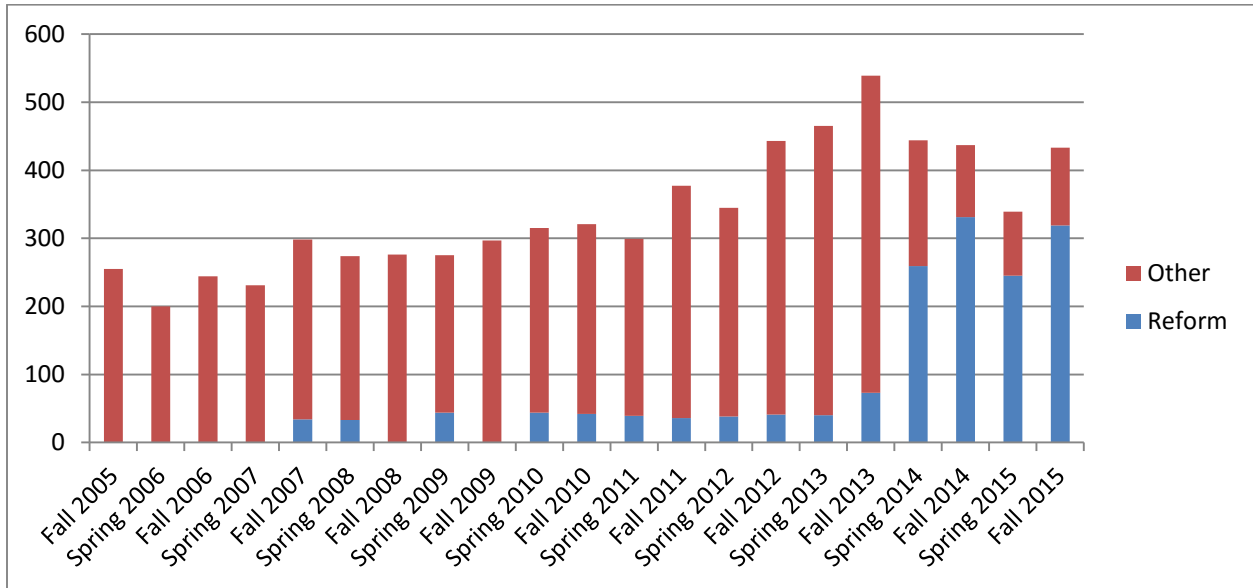

Figure 2: Calculus I enrollment by semester.

Total students "captured" by the reform project, as a percent of enrollment is shown in Figure 3. It appears to be stabilizing in the low to mid 70's, which currently reflects the portion of calculus that Boise State University has chosen to offer as honors, online, or face-to-face but parallel to online.
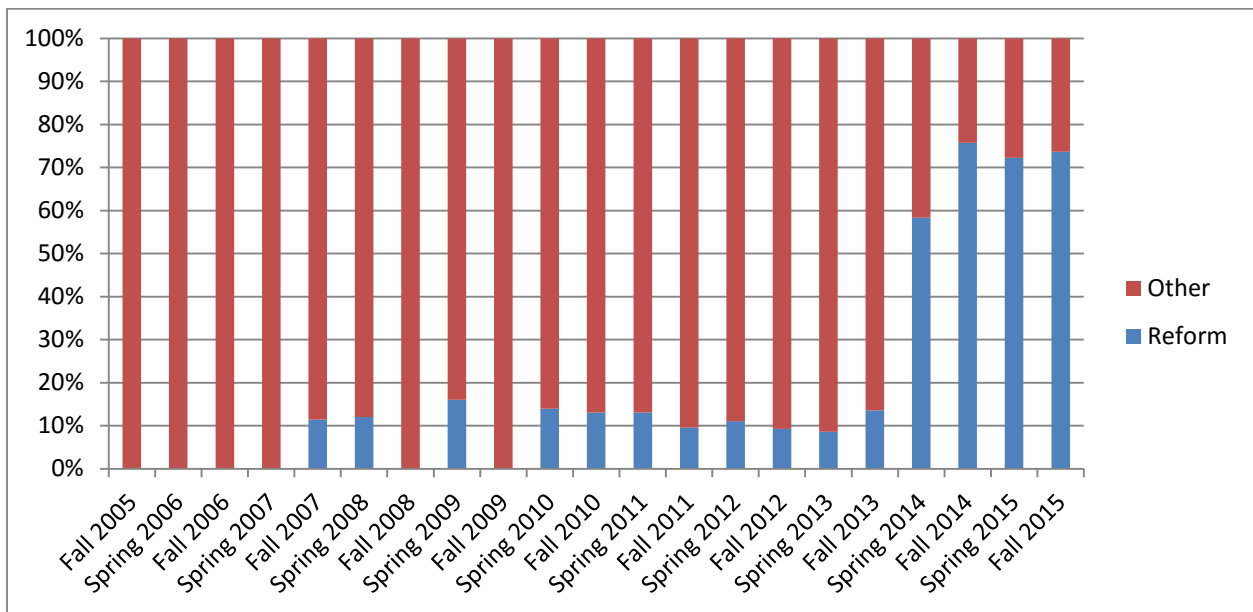

Figure 3: Enrollment in reformed Calculus I expressed as a percentage of enrollment.

The focus of this paper

Pedagogical reforms are subject to criticism for many reasons and from many quarters. As a leader or change agent, one must be prepared to address such concerns. This paper presents a rigorous, data driven technique for refuting such. Resistance takes the form of critiques such as "pass rate gains are probably grade inflation," or "important content is missing or mishandled in the new Calculus I". Any such claim is a testable hypothesis. If the claim is that the inflated grades or missing content result in harm to Calculus II students, then this is a claim that the null hypothesis, "Students do equally well in Calculus II, regardless of the Calculus I reform" is rejected in favor of the alternative hypothesis, "Students coming from reform Calc I do worse in Calculus II."

This paper reports on a natural experiment that provides a rigorous statistical test of the above hypothesis. We find that you cannot reject the null hypothesis, which means Calculus I reform causes no damage to students in Calculus II.

Additional resistance or continued skepticism is entirely possible. In particular, one cannot conduct this experiment without making many choices about how to select and interpret data. In anticipation of such we will explore several alternative choices. In this process we will sometimes see negative treatment effects, but never at any statistically significant level and never sufficient to reject the null hypothesis, even when the data are carefully chosen to make the treatment look as bad as possible. We will also see some positive treatment effects, meaning there are subsets of the data in which the treatment group out performs the control group in Calculus II.

Effects of Reformed Calculus I on Calculus II

To analyze student performance in Calculus II, we created two cohorts, further sliced by term, defined as follows:

Term X Treatment Cohort consists of all students who passed reformed Calculus I during term X at Boise State University, and then immediately enrolled in Boise State University Calculus II.

Term X Control Cohort is all students who passed non-reformed Calculus I during term X at Boise State University, and then immediately enrolled in Boise State University Calculus II.

We define "immediately" to mean without skipping a regular term, so spring-to-fall succession is immediate. This means that a spring cohort will contain students who took Calculus II in either the subsequent summer or fall. No student can be in more than one cohort. We capture only the first instance of a student passing Calculus I, and then only the immediate enrollment in Calculus II. We care only about passers of Calculus I because our dependent variables are performance metrics in Calculus II, which requires Calculus I as a prerequisite. We wish to compare the efficacy of reformed to non-reformed Calculus I, so we do not consider transfer Calculus I credits or other routes into Calculus II (CLEP, AP, etc.).

Cohorts are defined by term so that we can conduct time series analysis. However, the analysis in this paper will primarily deal with aggregated cohorts. Unless specifically stated otherwise,

both the treatment and control cohorts will be aggregated from spring 2013 to summer 2015, inclusive. We refer to this as **the baseline cohort**.

For each cohort, we measure two dependent variables: Calculus II pass rate and average Calculus II grade.[ii]

Calc II Pass Rate = (Number of A's, B's, C's) / (Cohort size – Audits – Incompletes)

Grades are converted to grade points on the usual 4-point scale. Boise State University uses +/- grades, so the conversion is A+ = 4.0, A = 4.0, A- = 3.7, B+ = 3.3, etc. We count W's and CW's as 0.0, along with F's. This is consistent with DWF used more generally as a student success metric, in that it considers an F and a W to be equally unsuccessful.

Calc II Grade = (Total grade points) / (Cohort size – Audits – Incompletes)

It is possible that there one cohortcould be better prepared for college level work than another cohort. So we tracked four control variables for each cohort:

GenACT: About 70% of our students have either an ACT or an SAT Math score. SAT Math scores are converted to ACT using published concordances.[7] If this results in two scores (some students have both ACT and SAT) we take the higher. This is averaged across members of the cohort that have at least one score.

CumGPA: This is the cumulative grade point average for each student at the conclusion of the cohort term, averaged across the cohort. Data limitations at Boise State University force us to use a value of Cum GPA that may have become slightly inflated by grade replacement in the semesters since the cohort term. The effect is small, but nonetheless unfortunate. In a future paper we hope to replace this variable with a more rigorously controlled GPA.[6] We continue to use this variable because it is the only independent variable for which we have data for all students, and because it is our only independent variable that directly measures ability to do college level work.

HSGPA: High school GPA. Along with SAT/ACT scores this is recognized as one of the strongest predictors of college success. We have data for approximately 85% of our students.

AdIndex: Boise State University calculates an Admission Index based on HS GPA and composite ACT/SAT. Also recognized as a strong predictor of college success. Since it uses composite test scores rather than just Math, it is not redundant. We have data for approximately 55% of our students.

[ii] "Dependent" in this context means only dependent on control vs treatment. Although there are also "independent" variables, we will not conduct regression analysis in this paper.

The Data

Here is the complete time series for all variables and both cohorts:

Variable	Cohort									
Cohort Size	Control	156	46	173	72	38	48	34	31	**598**
	Treatment	14		30	114		162	100	25	**445**
Calc II Pass Rate	Control	76%	67%	63%	61%	63%	71%	65%	48%	**66%**
	Treatment	57%		33%	61%		74%	66%	48%	**64%**
Calc II Grade	Control	2.35	2.16	1.88	1.71	1.75	2.20	1.89	1.56	**2.01**
	Treatment	1.95		1.04	1.76		2.27	1.93	1.67	**1.94**
CumGPA	Control	3.16	3.23	3.20	3.04	3.03	3.19	3.17	3.20	**3.16**
	Treatment	3.41		3.19	3.06		3.13	3.13	3.16	**3.13**
GenACT	Control	24.7	22.3	25.6	22.7	24.9	24.3	25.0	23.4	**24.6**
	Treatment	23.5		24.2	24.0		25.4	24.4	25.7	**24.7**
AdIndex	Control	58.2	49.2	64.6	58.5	54.7	65.7	64.4	53.7	**60.9**
	Treatment	64.0		56.5	58.6		63.7	59.2	55.5	**60.9**
HSGPA	Control	3.38	3.29	3.45	3.31	3.26	3.41	3.43	3.24	**3.38**
	Treatment	3.49		3.24	3.30		3.40	3.38	3.39	**3.36**

Table 2: Time series for all variables and both cohorts, spring 2013 – summer 2015.

The scale up semester, spring 2014, is highlighted in Table 2. The last column shows the aggregate measures. Statistical analysis follows, but we have found that this or similar charts are useful as descriptive data. Some observations:

- In aggregate, the treatment cohort scores slightly worse on both measures of Calc II performance. (Neither group is doing all that well in Calculus II, but that is a separate topic.)
- The four independent variables suggest, in aggregate, that the difference in academic preparation between the treatment and control cohorts is negligible. It is difficult to judge whether a 0.1 gap in average ACT Math score matters compared to a .03 gap in GPA, or a 2% loss in pass rate. This is what statistical tools are for.

Although the aggregate data looks at first glance worse for the treatment group, the time series from the scale up term onward actually shows treatment groups did slightly better in Calculus II. Not by much, but consistently and on both measures.

Analysis

We apply standard statistical tests to compare Calc II Pass Rate and Calc II Grade for each cohort.

We also conducted statistical tests to compare the average values of the independent variables for each cohort. For all independent variables, the hypothesis we are testing is whether the cohorts are actually any different. The results are summarized in Table 3.

Table 3: Effects and *p*-values for baseline cohort				
Values	**Control**	**Treatment**	**Effect Size**	***p*-Value**
Cohort Size	598	445		
Calc II Pass Rate	66.4%	64.3%	-2.1%	0.239
Calc II Grade	2.01	1.94	-0.07	0.215
CumGPA	3.16	3.13	-0.03	0.836
GenACT	24.60	24.70	0.11	0.359
AdIndex	60.88	60.91	0.02	0.493
HSGPA	3.38	3.36	-0.02	0.671

The effect size column recaps what was observed in Table 2. We see a small negative effect of treatment. However, the *p*-values are large, meaning the effect is of no significance. We cannot reject the null hypothesis, so we conclude that Calc II pass rates and grades for the two cohorts are not meaningfully different. The *p*-values for independent variables show that the two cohorts are not meaningfully different on *a priori* academic measures.

Discussion

The overarching purpose of this analysis is to provide change agents, campus leaders, and curriculum reformers the ability to present a persuasive and rigorous argument to potential resisters. Resisters may legitimately claim that these statistical results are influenced by choices made by the study designers, such as which terms to aggregate, whether to consider only immediate follow-on to Calculus II, etc. It is true that results are sensitive to such choices. Table 4 presents the results from aggregating only the three regular semester cohorts since scale up began.

Table 4: Baseline cohort restricted to regular terms from scale up to present				
Values	**Control**	**Treatment**	**Effect Size**	***p*-Value**
Cohort Size	154	376		
Calc II Pass Rate	64.9%	68.1%	3.2%	0.756
Calc II Grade	1.90	2.03	0.13	0.839
CumGPA	3.12	3.11	-0.01	0.575
GenACT	23.69	24.77	1.08	0.010
AdIndex	62.33	61.24	-1.09	0.695
HSGPA	3.36	3.37	0.00	0.492

This choice converted negative treatment effects into positive effects. However, it also created a larger and much more significant gap between the cohorts' ACT Math scores. This *p*-value is small enough to suggest that for this subpopulation more analysis, with some device to control for effects of this variable might be appropriate. However, for the purpose at hand, it is enough to note that the effects of treatment are non-negative. It is not possible for any positive treatment effect to lead to rejection of null hypothesis. What the finer analysis is really saying is that it would be a mistake to attribute the positive effects to the treatment itself.

Since the goal is to provide good evidence that Calculus I changes do no harm in Calculus II, , the best approach is to rerun the analysis with as many different reasonable modifications as possible. If the null hypothesis is true, then it will be unlikely that any modification results in data that calls for rejecting the null hypothesis. Moreover, if you have looked at the data in many ways, as we have, you can present as the baseline case a viewpoint that is least favorable to your project. Then repeated attempts to probe your data will mostly show better results for your project.

We have done this for the data set here. Some modifications that we have studied:

- Start at the scale up term, as in Table 4, but remove summers. This makes the treatment cohort look better than it does in the baseline case of Table 3.
- Put the summers back. Now the treatment cohort looks a lot better. In fact it is almost good enough (p =0.06) to reject the null hypothesis in the other direction, suggesting that the non-reformed Calc I is hurting students in Calc II.
- Restrict to subsets of students for whom we have complete data on independent variables. For example, Table 4 could be modified to look only at students who actually have ACT Math scores. (Oddly, the positive Calc II effects vanish – so they were not after all caused by the ACT scores. This sort of odd artifact is not uncommon when studying data sets in which the measured effects are statistically insignificant.)
- The control group contains all of the honors calculus sections. We could remove them. Unsurprisingly, this makes the treatment cohort look better.
- We included only students who took Calc II immediately after Calc I. There are good technical reasons for this choice, but someone could argue it was done to massage the data. To test this we added delayers to the data. This, too made the treatment cohort look better than before.
- Combinations of the above. All of them result in treatment results that are better than Table 3. Many of them result in positive treatments effects.
- At least one perfectly reasonable combination -- begin at the scale up term, keep all subsequent terms, and remove honors sections -- results in very attractive treatment effects: Table 5.

Table 5: Baseline cohort, from scale up to present, without honors				
Values	Control	Treatment	Effect Size	*p*-Value
Cohort Size	206	401		
Calc II Pass Rate	59.7%	66.8%	7.1%	0.957
Calc II Grade	1.72	2.00	0.29	0.994
CumGPA	3.07	3.11	0.04	0.202
GenACT	23.25	24.79	1.53	0.000
AdIndex	56.72	61.14	4.42	0.015
HSGPA	3.28	3.37	0.09	0.026

Here we see large positive effects of treatment. However, it is statistically irrelevant to the specific experiment and research question addressed in this paper. This is because we set out to

test a one-sided hypothesis – that the treatment cohort does no worse. If the observed result is that the treatment cohort does *better* then you cannot reject the null hypothesis. What this means is that, if we had chosen to design an experiment to test a different hypothesis – say that treatment improves Calc II outcomes, there is a good chance that we would have a positive result. We will focus on this in future work.

The larger point is that the decisions defining the baseline cohort are probably sound decisions, and are certainly defensible against any claim of massaging data in favor of positive treatment outcomes.

Here is a summary of the reasons for originally settling on our choice of baseline cohort.

- The time series from spring 2013 to summer 2015 is well balanced temporally. It includes 2 regular semesters prior to scale-up, during which control cohorts were larger than treatment cohorts. It includes 2 regular semesters after scale-up, when treatment cohorts were larger than controls. And it includes the scale-up semester itself, in which the cohorts were most nearly equal in size. Also, this time series achieves a nice middle ground between the desire to have large overall N, and to have N split into roughly equal sized treatment and control cohorts. (Older start dates clearly increase N, but older terms are also overwhelmingly control cohort. More recent start dates, like spring 2014 both lower N and skew the cohort size towards treatment.)

- We chose a cohort definition that restricts us to students who take Calculus II immediately after passing Calculus I. This lowers N by about 10%. (Approximately 90% of students who ever continue on to Calculus II do so without delay). Including delayers would skew towards older cohorts, since only the older cohorts have had enough time to delay and then eventually take Calculus II. This in turn means that control cohorts will have more delayers than treatment cohorts. Delayers may have different personal characteristics that influence Calculus II performance, which would be disproportionately present in the control cohorts.

- We stop with summer 2015 because it is simply the most recent term in which we could have any longitudinal data for students continuing to Calculus II (fall 2015).

- We left honors students in the control cohort because, if ever in doubt, do not restrict data in ways that obviously promote your result.

Effect of treatment on female and underrepresented students

This section examines how women and under-represented minority students perform in Calculus II as a result of the reform of Calculus I. Table 6 presents results for the baseline cohort restricted to female students.

Table 6: Baseline cohort, women only.				
Values	Control	Treatment	Effect Size	*p*-Value
Cohort Size	108	87		
Calc II Pass Rate	67.6%	64.4%	-3.2%	0.318
Calc II Grade	2.06	2.08	0.01	0.528
CumGPA	3.29	3.26	-0.02	0.621
GenACT	24.27	24.83	0.56	0.187
AdIndex	66.22	66.60	0.38	0.447
HSGPA	3.53	3.60	0.07	0.147

We see the usual results: small effects in Calc II, none significant, and we cannot reject the null hypothesis. The treatment group here has an edge over controls in the independent variables, so if there were anything of interest in the Calc II effects it would be best to control for the effects of ACT and HSGPA.

This is the first slicing of the data that even hints at possibly weaker performance in Calc II. Calculus II course grades are up a tiny bit; pass rate is down a larger amount; but both *p*-values indicate this not significant. Control variables are split. Mostly this data says female treatment and control groups perform the same in Calculus II.

Table 7 presents the results for underrepresented minority students (URM). URM is defined here using our Data Warehouse IPEDS Ethnicity field. We would prefer to have more nuanced information, but working with this we exclude all students who are classified as White, Asian, Non-resident Alien, Two or More Races, or Unknown. This drops our *N* from over 1000 to 108.

Table 7: Baseline cohort, URM only.				
Values	Control	Treatment	Effect Size	*p*-Value
Cohort Size	52	56		
Calc II Pass Rate	53.8%	66.1%	12.2%	0.904
Calc II Grade	1.79	1.88	0.09	0.624
CumGPA	3.13	3.08	-0.05	0.673
GenACT	23.37	22.88	-0.49	0.701
AdIndex	56.19	57.74	1.54	0.361
HSGPA	3.36	3.36	0.00	0.485

Unsurprisingly, with such low *N*, the difference between pass rate and GPA between the treatment and control cohorts is not statistically significant. Nor are the differences in academic ability of the two groups. However, the pass rate effect is large, indicating that further study may show that treatment causes pass rate to go up for some groups. That the reformed Calculus I resulted in increased pass rate for URM students is not unexpected; the literature clearly shows that active learning strategies, such as is deployed in the reformed Calculus I (group work, etc.), have a positive influence for students who are part of an under-represented minority group. For example, Klingbeil's longitudinal study[3] showed that graduation rates were tripled for URM students who took their engineering-based introductory mathematics course (a hands-on, application focused course) compared with those who did not, and that URM students had higher

graduation rates than the control group.

Statistical Significance of Calculus I Effects

Our treatment is an intervention in Calculus I. It is not surprising that it shows no statistically significant effect in Calculus II. It was designed to achieve effects in Calculus I. It does. This is what statistically significant data look like:

Table 8: Calculus I Effects.				
Values	Control	Treatment	Effect Size	*p*-Value
Cohort Size	1540	994		
Calc I Pass Rate	67.3%	72.8%	5.5%	0.001
Calc I Grade	2.03	2.21	0.18	0.001
CumGPA	2.95	2.93	-0.02	0.774
AdIndex	59.78	59.73	-0.05	0.524
HSGPA	3.33	3.33	0.00	0.559

The treatment and control cohorts are indistinguishable on independent measures of academic ability. But the treatment effects of increased pass rate and grade are significant at the strongest levels of *p*-value used in any experimental studies.

Summary

Calculus I reform has produced sizable, sustainable, and statistically significant gains in Calculus I pass rates and grades. The course pair report is a rigorous, data driven response prepared to consider claims that student success in reformed Calculus I is the result of grade inflation or weakened standards. It also addresses claims that content in Calculus I cannot be altered for fear of degrading Calculus II performance. Our data persistently shows that reform Calculus I students do no worse in Calculus II than their peers who came through traditional Calculus I. This presentation is strongly resistant to any claims of data massaging, since nearly every adjustment makes the treatment look better than the baseline case we began with. No interpretation of the data comes anywhere close to rejection of the null hypothesis in favor of the alternative that Calc I reform harms future performance in Calc II. Both groups perform equally which should soundly refute any accusation that standards or content in the reformed Calculus I are any sort of danger.

These results are unsurprising when taken in light of the vast body of work done on active learning and its impact on STEM learning. Our reformed Calculus I contains active learning strategies including group work. The literature overwhelmingly supports the importance of group work/collaboration in terms of student persistence and success in major, due to their increased engagement with one another.[8,9] A recent metaanalysis of undergraduate STEM literature shows that active learning leads to increases in examination performance that would raise average grades by half a letter.[10] Our results show an increase in pass rates by approximately that level.

Future Work

As a result of numerous positive outcomes associated with Calculus I reform, the reform is now spreading into Calculus II, with a roll-out point of spring, 2016. Faculty perceptions seem to be generally positive from the Calculus I reform[1] and we will continue to monitor this. We will continue to rigorously analyze student performance by looking at course grade performance and post-requisite course performance. We plan to also begin to monitor student performance in certain engineering courses for which Calculus I or II are prerequisites (Statics and Dynamics).

Acknowledgments

References

1. Bullock, D., & Callahan, J., & Shadle, S. E. (2015, June), "Coherent Calculus Course Design: Creating Faculty Buy-in for Student Success," Paper presented at 2015 ASEE Annual Conference and Exposition, Seattle, Washington. 10.18260/p.23694. https://peer.asee.org/23694

2. Callahan, J., & Schrader, C., & Ahlgren, A., & Bullock, D., & Ban, Y. (2009, June), The Implementation Of An Online Mathematics Placement Exam And Its Effects On Student Success In Precalculus And Calculus Paper presented at 2009 Annual Conference & Exposition, Austin, Texas. https://peer.asee.org/5541

3. Klingbeil, N. W., & Bourne, A. (2013, June), A National Model for Engineering Mathematics Education: Longitudinal Impact at Wright State University Paper presented at 2013 ASEE Annual Conference, Atlanta, Georgia. https://peer.asee.org/19090

4. Bressoud, D., M. Carlson, J.M. Pearson and C. Rasmussen. 2012. Preliminary Results of the Study of Characteristics of Successful Program in College Calculus. 12th International Congress on Mathematics Education. Seoul, Korea, July 8-15 (2012).

5. Callahan, J., Bullock, D., and Chyung, S. Y. (2012, June), "Both Sides of the Equation: Learner and Teacher," Paper presented at 2012 ASEE Annual Conference, San Antonio, Texas. https://peer.asee.org/21025

6. Mesa, V., Jaquette, O., & Finelli, C. J. (2009). Measuring the Impact of an Individual Course on Students' Success. Journal Of Engineering Education, 98(4), 349-359.

7. http://www.act.org/aap/concordance/pdf/reference.pdf downloaded 1/31/2016.

8. Kogan, M. & Laursen, M.L. (2014) Assessing Long-Term Effects of Inquiry-Based Learning: A Case Study from College Mathematics. *Innovative Higher Education*, 39(3), 183-199

9. Springer, L., Stanne, M.E., Donovan, S.S. (1999) Effects of small-group learning on undergraduates in science, mathematics, engineering and technology: A metaanalysis. *Review of Educational Research.* 69(1), 21-51.

10. S. Freeman, S.L. Eddy, M. McDonough, M.K. Smith, N. Okoroafor, H. Jordt and M.P. Wenderoth, PNAS 8410-8415, doi: 10.1073/pnas.1319030111.