An Experiment in "Flipped" Teaching in Freshman Calculus

Laura Anderson, Joseph Brennan

July 9, 2014

In the fall of 2013 the Binghamton University Department of Mathematical Sciences undertook an experiment in "flipped teaching" in Calculus 1. We wanted to compare flipped teaching and our traditional methods in several respects:

1. Which method leads to better student performance in terms of computational ability? In terms of more in-depth problem-solving ability? In terms of conceptual understanding? In terms of performance in Calc 2?
2. Were the answers to the above questions different for students coming in with weaker math skills than the norm? Were they different for people seeing Calculus for the first time than for those with a high-school calculus course behind them?

1 Background

1.1 Course setup

For many years Calculus 1 at Binghamton has taught in sections of about 30-40 students, with graduate student instructors teaching most sections. A single faculty member sets the syllabus and schedule and makes common exams. While individual instructors occasionally made small experiments in their teaching methods, this has always been a traditional lecture course. Calculus is where many instructors have their first experience teaching their own class. The faculty makes an effort to mentor new instructors, and while members of the department often discuss good traditional teaching, there has been little to no emphasis on innovation in teaching methods.

1.2 Previous success with technology

In the fall of 2012 we moved a portion of our testing online. Basic computational questions were moved from paper tests to proctored, computer-based "Skills Tests". Students had multiple tries to pass each Skills Test (with a new randomized test on each attempt), and students were required to pass all Skills Tests in order to pass the class. Paper tests were then reserved for more sophisticated or conceptual problems.

While no thorough analysis was made of this change's impact, two outcomes were obvious:

1. Student reaction has ranged from neutral to vehemently negative. The inflexible bar of having to pass these exams and the lack of partial credit greatly concerned some students.

2. Students' performance and grades improved markedly. While we endeavored to maintain the same grade standards from past semesters, in the first semester of the Skills Tests the percentage of students getting grades of D, F, or Withdraw dropped from 24% to 19%, and the percentage of students getting A's rose from 22% to 28%.

This success whetted our appetite for further experimentation.

1.3 Motivations for experimenting with flipping

Some of our motivations for embarking on this experiment:

1. Two of the great challenges in Calculus 1 are maintaining student engagement and addressing the needs of students with diverse mathematical backgrounds.

2. James Pittaresi, a Mechanical Engineering professor (now director of Binghamton's Center for Learning and Teaching) has had great success with flipped teaching in his sophomore mechanical engineering course. He advocated for the experiment and offered his guidance.

3. Enrollment at Binghamton University has been growing, and in recent fall semesters the number of sections had climbed into the 30's. Providing meaningful mentoring to new instructors has become more difficult, as has maintaining uniformity across sections. We hoped that moving some of the content delivery to videos would address uniformity issues, and creating a class format focused on interaction would show instructors how to engage with students.

A pressing question for us was one of scale. Other experiments we were aware of in flipped teaching at the college level were either by individual instructors or by small groups of experienced professors. Our eventual aim was a framework for flipped teaching that could be taught to and carried out by our many relatively inexperienced graduate student instructors.

2 Experimental setup

Our essential plan was to divide a relatively homogeneous group of students randomly into two groups, one of which would would be taught in a flipped format, while the other was taught in a traditional format. The textbook, course schedule, etc., was the same. We would then compare the outcomes for the two groups by several metrics:

1. the computer-based Skills Tests on basic computational skills,

2. two paper-based midterms on more sophisticated problem-solving or more conceptual problems,

3. a common final exam with both types of problems,

4. a common final exam from Calculus 2*final exam grade, or overall performance in Calc 2?*, and

5. the Calculus Concept Inventory (described below).

Eleven sections of Calculus 1 are reserved for Engineering students. These are all first-term freshmen carrying similar academic loads. These 11 sections meet three times per week at 8am for 90 minutes, and students are assigned to sections by a staff member in Engineering – they do not choose their own sections. Thus they formed a good cohort for our plan.

This group of students, however, varied greatly in their mathematical preparation. Independently of this experiment, we had instructors struggling with the wide range of student abilities, and we wanted to separate out students with weak precalculus skills into separate sections tailored to their needs. Thus we created four flavors of section:

1. one section of weak students (as measured by our precalculus Screening Test) taught in a traditional format
2. two sections of weak students taught in a flipped format
3. five sections of typical students taught in traditional, and
4. three sections of typical students taught in a flipped format.

Engineering students were assigned to a flipped or traditional section randomly and, initially, without their knowledge. (They were told in advance whether they were assigned to a weak section or typical section.)

Students were given a questionnaire at the beginning of the course and had the option to identify their calculus experience. Not all students completed the questionnaire.

	Previous Calculus		Previous Precalculus		
	Yes	No	Yes	No	SAT Math Average
Flipped	104	18	102	20	691
Control	136	54	182	8	663

In the preceding spring, we sent a mass email to all of our funded graduate students soliciting volunteer flipped teaching instructors. We had about twice as many volunteers as we had spots for, and we endeavored to select a diverse group, including both inexperienced and highly experienced instructors, men and women, quiet and outgoing, and native and non-native English speakers. We then tried to match these with comparable instructors for non-flipped sections.

The Calculus Concept Inventory, developed by [**?**], is a test of conceptual understanding of basic principles of calculus. It is widely used in experiments similar to ours. It is a multiple-choice test involving no computation. We administered the test at the beginning and the end of the semester, and for each section we calculated the *normalized gain*, defined to be

$$\langle g \rangle = \frac{\mu_f - \mu_0}{100 - \mu_0}$$

where μ_0 is the mean percentage score of the class at the beginning of the semester and μ_f is the mean percentage score of the class at the end of the semester.

3 Development and execution

As our colleague, Dr. Pittaresi, was a large motivating factor in our experiment, we modeled our format on his Statics and Dynamics course.We would consider our "flipping" to be light, with a typical flipped class meeting consisting of

1. one or two short (< 10 minutes) videos to be watched in advance,
2. a short online homework to be done in advance, testing only understanding of the basic material in the video,
3. the actual class meeting, in which instructors interspersed short spells of lecture and class discussion with group work on a series of problems.

Our first step was to go through the textbook section-by-section and to decide which lent themselves to this format. We decided that most sections of the Calculus 1 curriculum lent themselves to "flipping"; a 5-10 minutes video could convey the basic concepts, and then the details could be developed in class through examples worked by the students. We decided to not "flip" our introduction to continuity/differentiability, integration with Riemann Sums, and cross-sectional volume.

Our second step was to develop the video and the in-class work for each class. Much of this was done in the summer, and it was time-consuming! After browsing the internet, we decided to develop our own videos and tailor them to our student population. The videos were made using Camtasia on a Microsoft Surface tablet computer – the lecturer's speech was recorded together with the content of a screen, on which the lecturer wrote. The lecturer did not make an appearance in any video. The two of us had intended to both record lectures, but here we hit our first major obstacle – one of us (Anderson) found the technology maddeningly unwieldy, and she regretfully ceded the task to Brennan. Brennan, on the other hand, much enjoyed the process and became skilled at creating polished screen-casts. We recommend to anyone considering recording lectures to experiment personally with the hardware and software possibilities before committing to anything. On average, a 10 minute video took four hours to develop a script, generate graphics, shoot, and edit.

The in-class work consisted of a sequence of problems, organized into short lectures separated by problem solving sessions. Our activities draw heavily from Good Questions at Cornell and betterfilingcabinet.com and generally consist of

- Video Review Questions
- Conceptual Multiple Choice
- Computation (easy to hard)
- Conceptual Free Response

4 Informal observations

All but one of the flipped section instructors were extremely enthusiastic about what they saw in class. Student engagement ran high throughout the 90-minute classes, students of many different ability levels were demonstrating gratifying insight in class discussion, and instructors felt they were able to address the needs of individual students well. Additionally, daily attendance was consistently high and, thanks to the interactive classroom and caffeine, it was exceedingly rare for a student to doze off; recall that all sections were scheduled for $8:00$am.

Being newbies to this, we were unable to give instructors experienced-based training before sending them into the classroom. One flipped section instructor, who had previously been successful in traditional lecture courses, found himself to be very uncomfortable with the format, and despite his best efforts his students did worse than other sections. We believe careful and extensive instructor training is necessary to scale the flipped format up to classes with many instructors.

Initially, many of the students in both sections were not thrilled with the idea of "flipping" a classroom. In fact, the faculty coordinator heard from irate parents upset and confused with our pedagogy, believing the course to be fully online.

Many students in the flipped sections reserved for weaker students were very resistant to the teaching style. Many of these students lacked the confidence or the mathematical foundation necessary to become self-sufficient problem solvers. They responded much better when the in-class problems were rearranged to give more basic computational problems at the beginning of the class. Another great success with this cohort was to get them working in carefully selected groups, placing the weakest students with the most extrovert (regardless of ability) students.

On the other hand, students in the flipped sections reserved for strong sections embraced the teaching style. As they typically had confidence and a good mathematical foundation, these students quickly mastered basic computational problems and, as a consequence, more time was devoted to deeper conceptual problems.

Comments from a graduate instructor of a flipped strong section:

"Student response to the flipped classroom varied widely, but it was rarely negative. Often students engaged with the material more voraciously, and the results paid off. The flipped method places the instructor in the role of facilitator, rather than lecturer. That has a number of consequences. It encourages students to think on their own, be critical, and believe in and encourage their own abilities, rather than focusing more on their instructor. It forces them to take risks which, at first uncomfortable, become more natural and from which they learn greatly. These are the skills we want to be teaching students for life, not just mathematics.

Though truly a student centered approach, practice of the flipped method also helps good teachers become great teachers. The value of seeing how students think about and solve problems in real time is irreplaceable, and it drives the way I think about and teach material now regardless of what techniques I'm using. It also gave me daily experience handling students in an extremely dynamic environment, responding to questions which were interesting, valuable, and all the more potent because they were fueled by student curiosity.

Another result was a change in my teaching reviews. They were more balanced, expressing the drawbacks as well as benefits, and were overall still positive. In effect, students felt more comfortable expressing themselves and sharing their critical thinking, just as they were instructed to do in class on a daily basis. This experience was confirmed by a couple other professors."

5 Quantitative results

We collected quantitative data through four means:

- Skills Tests;
- Common Midterm and Final Examinations;
- Calculus Concepts Inventory;
- Common Calculus 2 Final Examination.

5.1 Skills Tests

The Skills Tests are intended to set a firm lower bar on passing the course. Students are given a thorough selection of practice problems in advance, they have multiple tries on each test, and the problems selected for these tests are very basic. In addition, students are highly motivated to pass these tests – they must get a score of at least 70% on each Skills Test in order to pass the class. Thus most students do pass – in fact, they typically eventually get a good grade. This limits the effectiveness of Skills Tests grades as a comparison of flipped vs. traditional format sections:

Flipped Sections

	ST 1	**ST 2**	**ST 3**	**ST 4**
A	72%	59%	47%	66%
B	23%	23%	19%	17%
C	6%	16%	33%	16%
F	0%	2.3%	1.6%	0.8%

Controlled Sections

	ST 1	**ST 2**	**ST 3**	**ST 4**
A	71%	55%	51%	69%
B	19%	25%	15%	16%
C	10%	17%	33%	14%
F	1%	3%	2%	1%

At the very least, we can say that our flipped teaching method did not impair students' basic computational skills.

5.2 Midterm and Final Examinations

Midterm examinations were split into two categories: computation based Basic Skills Tests and concept based Midterm Exams. The flipped and controlled sections took common Midterm Exams, but these exams were unique to sections which met at 8am so there is not a coursewide comparison. With a P-value of 40.4, the difference between the flipped and controlled Midterm Exam mean is **not** statistically significant.

Common Midterm Exams

	Flipped	Controlled
Average	64.8	62.9
3 Quartile	77.5	76
Median	67.5	65
1 Quartile	54	51.5
SDeviation	17.84	18.1

The final exam is administered simultaneously coursewide and is where computation and concepts are reunited. Students in the flipped sections did particularly well in comparison to students in the controlled sections and students coursewide. With a P-value of 0.085, the difference between the flipped sections and students coursewide is very statistically significant. Similarly, with a P-value of 0.002, the difference between the flipped sections and controlled sections is very statistically significant.

Final Exam

	Flipped	Controlled	Coursewide
Average	69.1	61	63.9
3 Quartile	79.5	74.88	76.5
Median	70.75	62.5	66.5
1 Quartile	60.25	49.13	53
SDeviation	14.6	18.1	17.6
Number	126	186	896

The comparison between strong students shows greater evidence that students benefit from a "flipped" format. With a P-value of 0.0007, the mean final exam difference between the strong flipped sections and strong controlled sections is very statistically significant. Notably, the first quartile of the strong flipped sections equals the coursewide average. Further, the mean of the strong Flipped sections was 40% of the coursewide standard deviation above the coursewide mean.

Final Exam - Strong Sections

	Flipped	Controlled	Coursewide
Average	71.8	62.2	63.9
3 Quartile	80.63	75.5	76.5
Median	73	64	66.5
1 Quartile	64	50.5	53
SDeviation	13.1	18.6	17.6
Number	96	156	896

Predictably, students who had previously had calculus in the strong controlled sections performed better than those that had not previously taken calculus. Notably, those students in the flipped sections who identified that they had not previously taken calculus outperformed those that had; however, with a P-value of 42.52 this difference is not statistically significant.

Final Exam - Strong Sections

Previous Calculus			**No** Previous Calculus	
Flipped	Controlled		Flipped	Controlled
71.9	62.8	Average	73.3	58.5
80.5	76.5	3 Quartile	81	73.5
73	66.5	Median	70.5	55.25
64	51.25	1 Quartile	66	46.5
12.0	18.7	SDeviation	12.3	18.2
76	103	Number	17	44

Though the weaker students in both flipped and controlled sections performed worse than their strong counterparts, in terms of raw scores the weaker flipped students outperformed the controlled students. However, with a P-value of 7.72, the difference is **not** statistically significant.

Final Exam - Weak Sections

	Flipped	Controlled
Average	60.3	54.7
3 Quartile	71.63	63
Median	60.5	55.5
1 Quartile	51.88	40.5
SDeviation	16	13.6
Number	30	29

We initially set out to follow students with the Watson College of Engineering, however we were only able to obtain college enrollment on students from the end of Spring 2014. With a P-value of 0.04, the difference in the final exam mean among Watson students is very statistically significant.

Final Exam - Engineers Only

	Flipped	Controlled
Average	68.1	59.6
3 Quartile	78	74.63
Median	70.5	60.25
1 Quartile	60.5	45.88
SDeviation	13.3	17.7
Number	83	68

5.3 Calculus Concepts Inventory

Of the students that took both exams, there was a 1.2 point and 1.3 point increase on average in both the flipped and controlled sections, respectively. The normalized gain for the flipped sections is 9.7% while the normalized gain for the controlled sections is 9.3%.

5.4 Success in Calc 2

For every pairing, the differences in the means between flipped, controlled, and coursewide sections is not statistically significant.

Common Final Exam - Calculus 2

	Flipped	Controlled	Coursewide
Average	63.4	60	64
3 Quartile	71.1	70	75.6
Median	64.4	60	64.4
1 Quartile	53.3	50.4	51.7
SDeviation	15.6	18.8	18.8
Number	75	63	340

6 Survey results

Student Satisfaction:

Videos	Activities	Overall	WebAssign	WebWork
3.36	3.18	3	2.64	2.55

- "I think it is fantastic. Seeing cautions whenever you need to review is very useful. The teacher is able to review any parts of the video the class did not understand and we can go over problems in class."
- "It is nice to get feedback and help on questions and have lots of homework. If anything is still unclear, we can bring it up to the instructor. More homework would actually be helpful."
- "I was in a lecture format at the beginning of the semester and this section is definitely preferable. It is nice to get feedback and help on questions and have lots of homework."
- "I like how class is for practice, but you learn as you practice. The videos provide basic background info. It has been interesting."
- "I don't want to take a class like this ever again. Next semester I hope to take a more traditional style of teaching calculus."
- "I dislike this format completely. I would learn much better if I was taught be an instructor that I could ask questions. It would be better if there were more lectures."
- "Sometimes, I would like to go over more of materials from videos in class, it is confusing. If the video is not understood, classroom activity becomes unproductive."

7 Summary

Stats: Unpaired t-test.

Note on SUNY CIT IITG Grant

- The Calculus Concept Inventory – measurement of the effect of teaching methodology in mathematics, Jerome Epstein, Notices of the AMS vol. 60, number 8

Maybe mention our exception from the IRB at Binghamton, no student decided to remove their data from our experiment???