Sources of variation in student success in Chem 111

Trevor Moore, Robin Donatello

Spring 2018

Introduction

To understand if a particular intervention or redesign is effective in improving student success, it is crucial to understand the recent trends in student characteristics and student success for the population of interest. In the current environment of pushing improvement in various performance metrics (such as DWF rate) it is critical to know what is the background or existing level of variation in these metrics. The goal is to inform the implementors of various student success projects about the effect of the intervention, but not over-stating these effects if they are small relative to the existing or historical background fluctuation or "noise".

This report also serves as an initial attempt to update a previous analysis on Sustaining Success in Chemistry (available online at https://norcalbiostat.github.io/chem_ss/). Chico State is in the middle of moving and revamping their entire data systems, so access to institutional data used in the previous analysis has been reset. At this point the data systems are still incomplete, and thus this report has limited scope in terms of student level characteristics. However, as will be seen it still serves as a source of data verification and reproducibility, along with highlighting some existing issues with the data and identifying current trends in student performance.

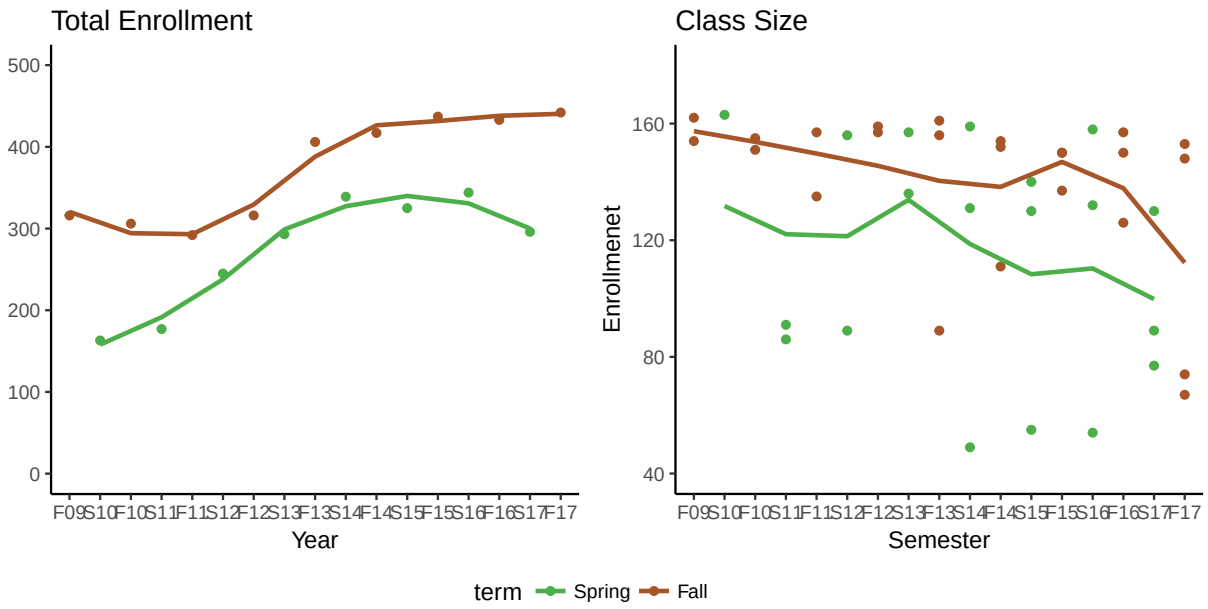
This report analyses the cohort of students taking Chemistry 111 between Fall 2009 and Fall 2017. Students withdrawing or receiving an incomplete (`W/I`) are excluded from this analysis. A repeatable grade is defined as a student receiving a `D, D+`, or `F` grades, or an unexcused withdrawal `WU`.

We start with an overview of the course characteristics for all sections of Chemistry 111, move to available student level characteristics, and conclude with some preliminary comparisons between student performance in courses with SI and without.

Course Characteristics

Enrollment and class size.

Enrollment in Chemistry 111 has steadily increased since 2009, with Fall semester seeing about 100 more students compared to Spring. Fall class sizes remained relatively constant, between 140-160, between 2009-14. An additional section in both Fall and Spring were added in AY 13-14. This additional section tends to have substantially fewer (sometimes less than half) students enrolled compared to the large lecture classes. The trends lines suggest that since Fall 2015 or so sizes appear to be decreasing, but this is more likely due to a lower enrollment in that additional section. The total number of students enrolled across all sections is either increasing or constant.



Course Performance

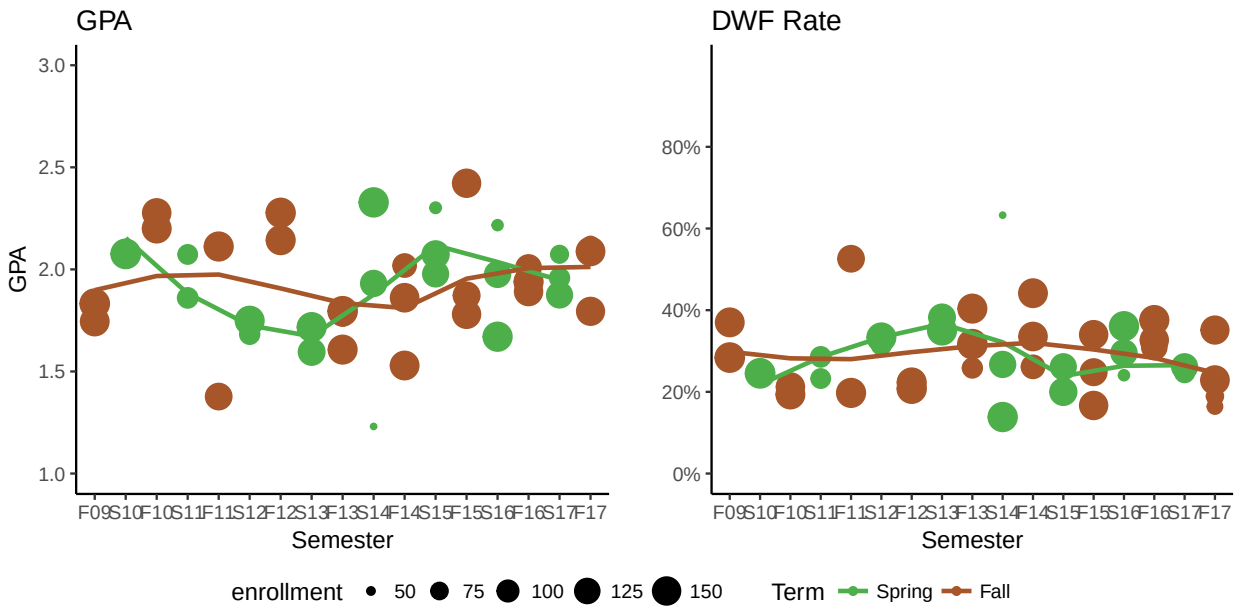
For the next two plots, the size of the dot represents the size of the class. This is an important characteristic since a DFW rate of 30% for a class of 100 is larger than for a class half that size.

There is no difference in the average GPA nor the DWF rate between fall and Spring semesters when pooled across all years under consideration (Table 1, p=0.45). However Fall does have marginally higher variance in GPA compared to Spring (p=0.047).

Table 1: Summary statistics for all years combined.

	N (%DFW)	Mean GPA	SD GPA
Spring	622 (28.5%)	1.91	1.11
Fall	989 (29.4%)	1.94	1.15

Spring and Fall semesters have different patterns of oscillating increase and decrease of class-level GPA, but there is no statistical linear trend across terms nor a difference in average GPA between Fall and Spring. As expected we see a similar pattern, slightly different between Fall and Spring, for the number and % of student receiving a repeatable grade.



Time of day

The common feeling between students and faculty alike is that 8am classes are "horrible", but does it really affect course performance? It turns out that noon has the lowest average GPA and the highest DFW rate, with 9am as the highest average GPA, but also the highest variance in GPA.

Table 2: Summary statistics for all years combined.

	N (%DFW)	Mean GPA	SD GPA
8am	125 (27%)	1.925	1.091
9am	508 (27.8%)	2.002	1.171
10am	180 (29.7%)	1.894	1.129
Noon	773 (30.8%)	1.872	1.126

Looking at the historical variation across semester, we see how GPA in the 9am class is the most variable, noon is a bit more consistent (lower variation), but 8am looks no worse than other times of day.

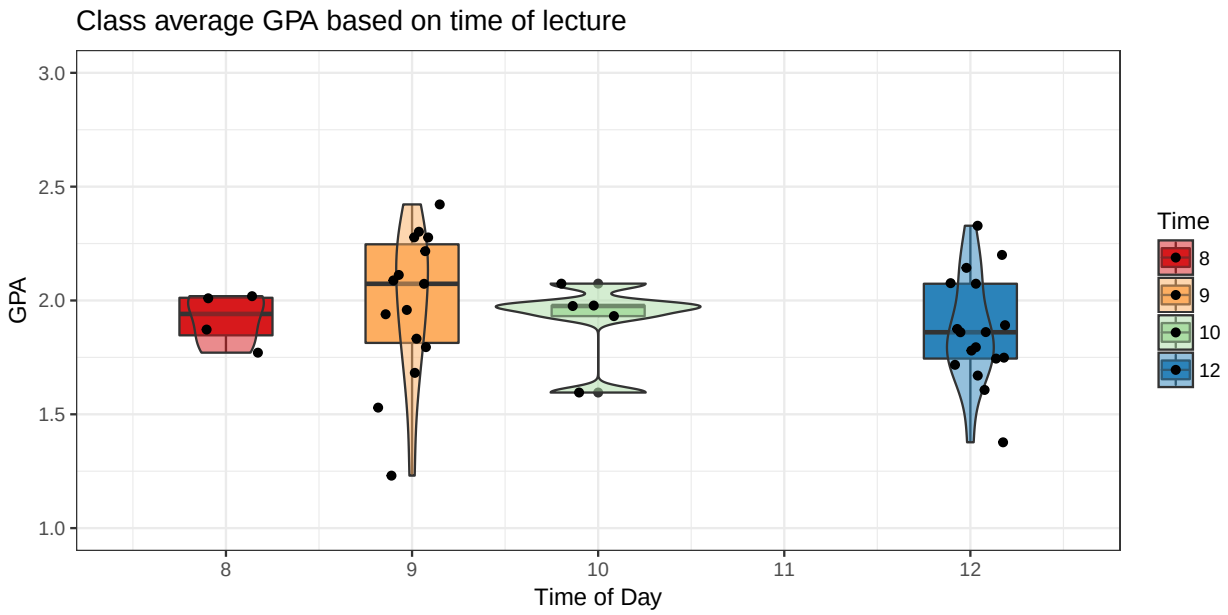


Figure 1: Average GPA across all sections per semester, by time of lecture

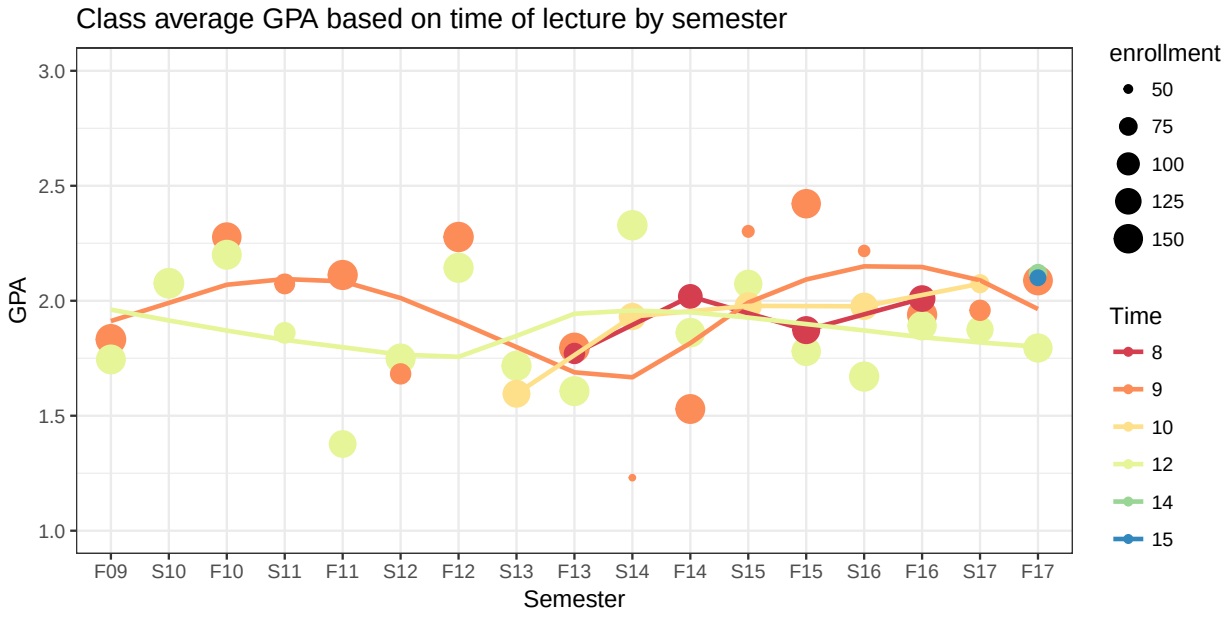


Figure 2: Average GPA across semester, by time of lecture

In Figure 1, the class-level average GPA exhibits large fluctuations in GPA, nearly up to an entire grade point. The 9 AM section shows the highest level of variation, with 12 Noon less variable due to a more consistent offering.

It is important to note that lectures starting in the afternoon (2pm, 3pm) were only offered in the Fall of 2017 (Figure 2). If these afternoon start times are offered in subsequent years following Fall of 2017, it may be valuable to examine the difference in GPA between classes in the morning and afternoon.

Student Characteristics

This section examines trends about student level characteristics over time.

Gender

Chemistry classes average about 54% male in the Fall and 60% male in the Spring, with the trends between fall and spring converging converging on slightly under 50% female. Things to consider that may be related to a lower proportion of females in a particular class include the time of day the class is being offered, and if the class was an additional section added at a later point due to a high wait list (or a high proportion of students needing to re-take the class).

Table 3: Sample size and proportion of males and female students for all years combined.

	Female: N (%)	Male: N (%)
Spring	875 (40.1%)	1306 (59.9%)
Fall	1546 (45.9%)	1817 (54%)

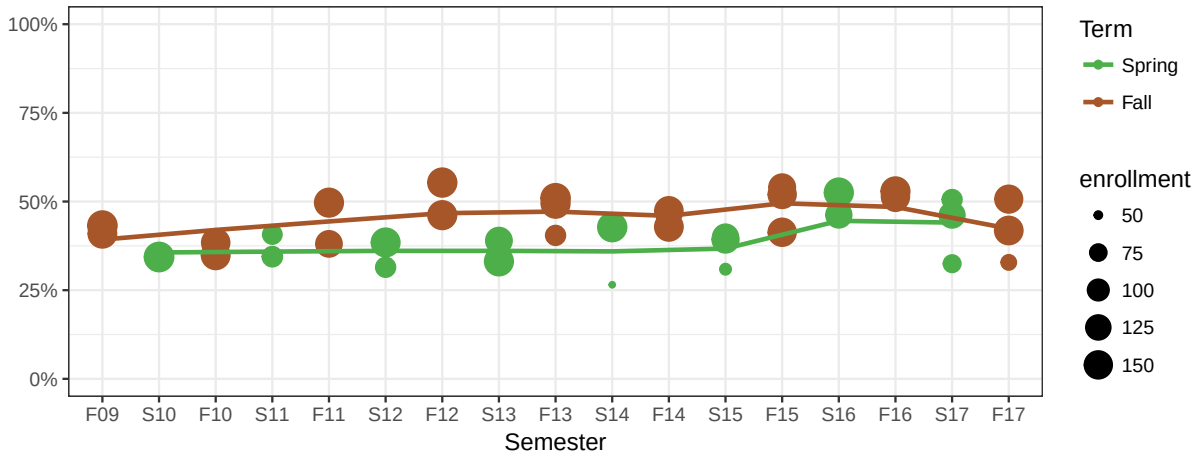


Figure 3: Proportion of female students across time by term.

Race/Ethnicity

Figure 4 describes the proportions of each ethnic category in a pie chart. It is of significant note that nearly half of the records in this data set have no value (N/V) for ethnicity. This is a known data issue that is an artifact of the current data migration into the new data warehouse.

This problem is made very clear in Figure 5 when we look at the proportion of students flagged as URM in the data systems since 2009. The definition used in this report follows the campus definition for an underrepresented minority student and is as follows:

- Non-URM: Asian, NHOPI, White
- URM: Black, HL
- Unknown: Mult, Unk

Figure 5 illustrates that ethnicity data prior to Fall 2014 is invalid. This also highlights that the proportion of individuals reporting multiple or unknown race/ethnicity is still in the neighborhood of 10%. Any binary categorization of URM that groups this "Unknown" URM category into either Non-URM or URM will have an effect on any comparisons. This makes it difficult if not impossible to draw valid conclusions utilizing URM status.

A better approach rather than dropping or combining this unknown group would be to cluster them individually into either URM or Non-URM based on which group they are more "similar" to using many other characteristics. Possible methods include propensity score matching and k-nearest neighbors.

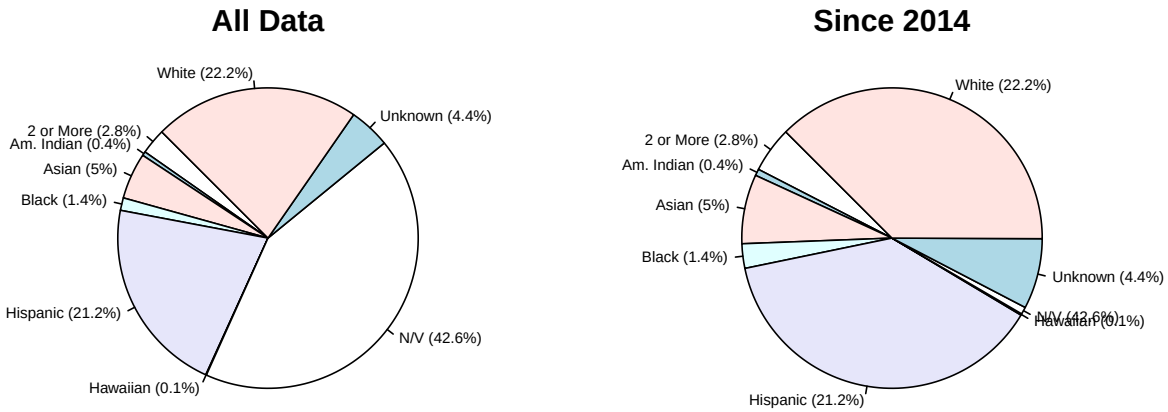


Figure 4: Distribution of Racial/Ethnic categories

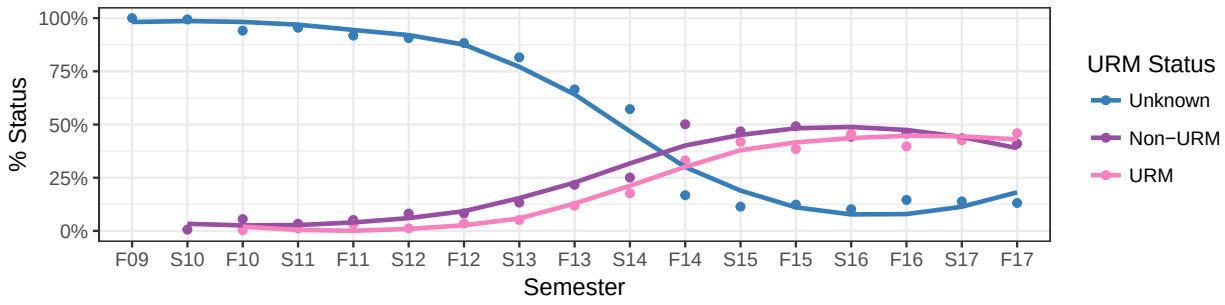


Figure 5: Proportion of individuals in each URM category over time.

Supplemental Instruction

SI was available to all Chem 111 students starting Spring 2016. Figure 6 illustrates that the average GPA after Fall 15 seems to be more stable (less varying) than in prior semesters. There also may be an upward trend in the average GPA across all Chem111 students since Spring 16, but these values still exist within the historical range.

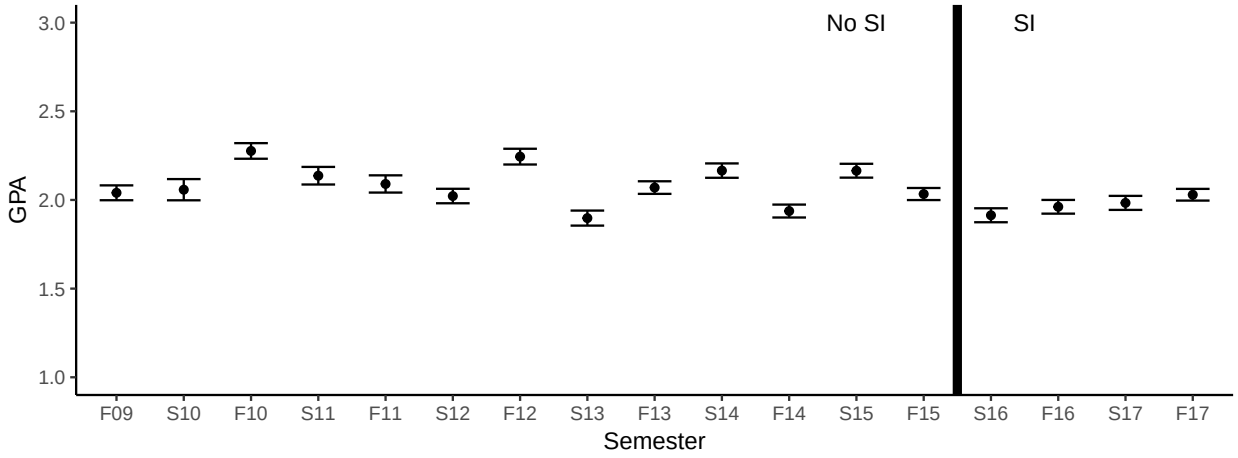


Figure 6: Average GPA (±1SE) across all students per semester

Next we examine the question: How does Supplemental Instruction modify measures of student success in CHEM 111 when considering various demographics?

Gender

In Figure 7, we consider the differences between GPA between genders. The average GPA for each semester fluctuates around 2.0 for both genders. However, there is no statistical linear trend across terms nor a difference in average GPA between males and females when considering all time points.

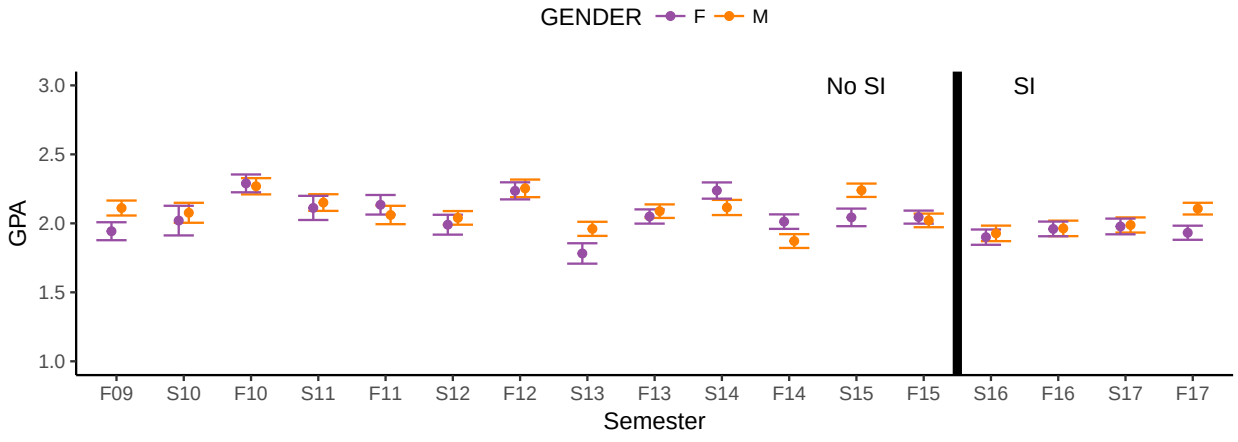


Figure 7: Average GPA (±1SE) across all students per semester, by gender

The average across all students will have a lower standard error compared to averages within sections. The variance of the average is always lower than the variance of the class, or the individual. Figure 8 shows how the gap in GPA between genders varies greatly by section (shapes).

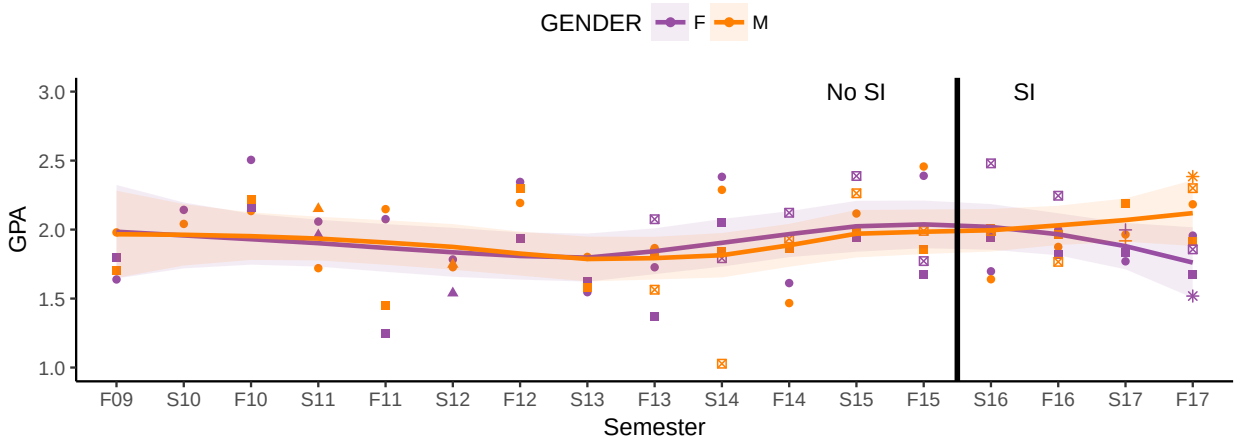


Figure 8: Section level differences in GPA by gender

Of interest to note, is the diverging trend in GPA between males and females after SI was introduced. The magnitude of this divergence is larger than what has been seen in the past 8 years. This could be a critical finding and thus warrants further examination.

Instead of looking at the absolute value of GPA between genders, we calculate the "equity" gap as average GPA for females - average GPA for males, within each section. Figure 9 The following plot then shows the trend in the variation in gap across time.

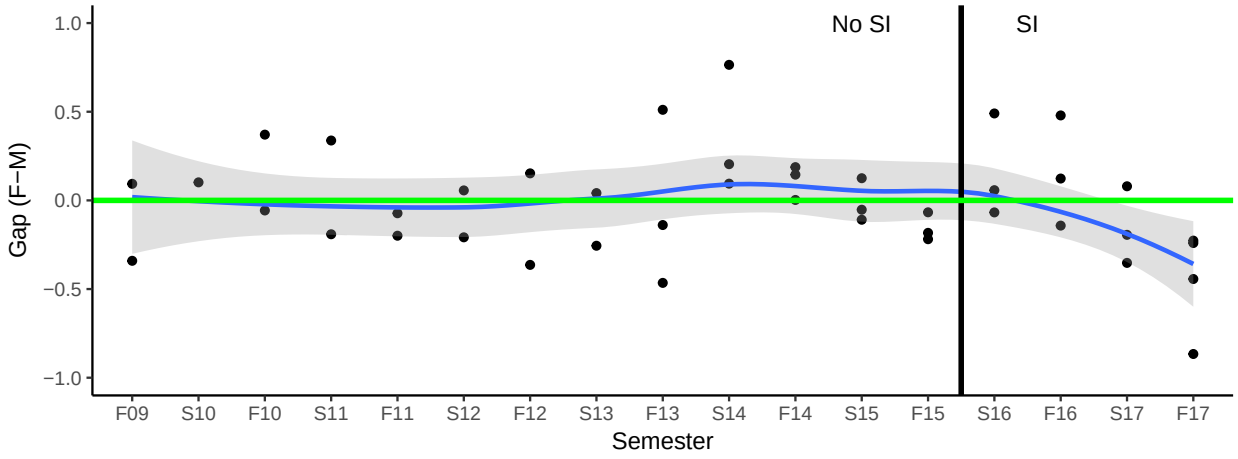


Figure 9: Class level equity gap between Females and Males within the same section

Across all sections, there is roughly a 0.3 difference in average GPA between males and females within the same section. This is represented by the grey band in Figure 9. Points above the green line indicate that females outperform males within that section, and conversely points below the green line indicate that males outperformed females within that section.

Here we see more clearly that in both Spring and Fall 17, males are tending to substantially outperform females in the same section. While one may hope that it may be because male's GPA is increasing at steeper

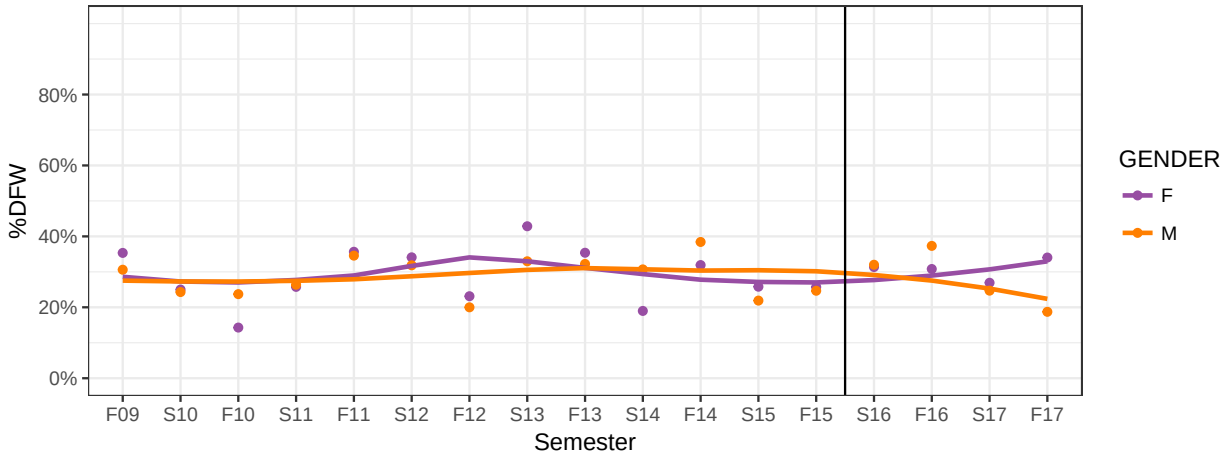


Figure 10: DWF rate by gender across semester

rate compared to females, Figure 8 indicates that the gap is increasing due to both an increase in GPA for males, but also a *decrease* in GPA for females.

Similar to GPA, the DFW rate oscillates between 20 to 40 percent for both genders. Additionally, it appears that there is no statistical linear trend across terms nor a difference in the DFW rate between males and females.

Statistical comparison between SI and non-SI

Here we considered only the classes from the spring of 2014 forward based upon the understanding that the differences due to shifting demographics in the past 10 years would be minimized. This allowed us to compare a similar cohort of students that were not offered supplemental instruction to those in the spring of 2016 forward, where supplemental instruction was available.

Figure 11 demonstrates that the distribution of GPA for this cohort is similar across spring vs fall semesters, gender, and SI status.

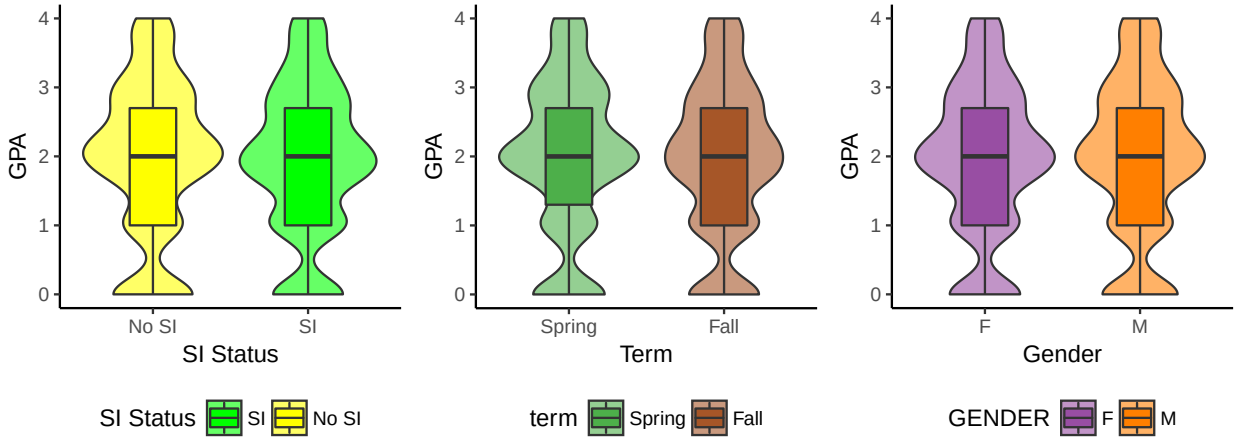


Figure 11: Distribution of GPA by SI status, Gender, and Term for the post 2014 cohort.

In figure 11, it appears that the mean values as well as the variance are very similar between cohorts offered SI and those that were not. This similarity between groups was confirmed using two sided T-tests to compare the mean GPA between groups.

Since student characteristics and performance is not bivariate in nature (not only due to one aspect) we fit a multivariable linear model of GPA and a logistic model of DFW rate using SI, term, urm status and gender as predictors.

Table 4:

	Dependent variable:
	GPA
SI vs Non-SI	0.01 (−0.07, 0.09)
Fall vs Spring	−0.05 (−0.12, 0.03)
M vs F	0.004 (−0.07, 0.08)
Non-URM vs Unknown	0.21*** (0.10, 0.32)
URM vs Unknown	−0.28*** (−0.39, −0.17)
Constant	1.99*** (1.88, 2.11)
Observations	3,033
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Odds ratios and 95% CI for receiving a DFW

	OR	CI	p
SI vs Non-SI	1.02	(0.87,1.2)	0.8193
Fall vs Spring	1.18	(1,1.39)	0.0512
M vs F	1.04	(0.88,1.22)	0.6734
Non-URM vs Unknown	0.76	(0.6,0.96)	0.0198
URM vs Unknown	1.64	(1.31,2.06)	<0.001
Intercept	0.32	(0.25,0.41)	<0.001

After controlling for other predictors, only URM status is a significant predictor of both GPA and DFW rate. URM students have 0.28 (.17, 0.39) lower GPA on average, and 1.6 (1.3, 2.1) times greater odds of receiving a repeatable grade compared to the 'unknown' ethnicity group. Non-URM students have 0.21(0.10, 0.32) higher GPA on average and 0.8 (0.6, 0.96) times the odds of getting a repeatable grade compared to those with 'unknown' race/ethnicity.

This is a demonstration that the group of students with 'unknown' race/ethnicity do not necessarily behave similar to either URM or non-URM groups and so should not be automatically grouped in with either of them as a whole.

SI as a moderator

Prior analysis indicates that SI may be a moderator in performance for different groups. Lastly we look at several interaction models where SI is allowed to moderate the relationship between the demographic and the outcome variable (GPA or DFW).

Table 6:

	Dependent variable:		
	GPA		
	(1)	(2)	(3)
SI vs Non-SI	−0.11* (−0.24, 0.01)	−0.06 (−0.18, 0.05)	0.04 (−0.15, 0.23)
Fall vs Spring	−0.15*** (−0.26, −0.04)	−0.05 (−0.13, 0.03)	−0.05 (−0.13, 0.03)
M vs F	0.001 (−0.08, 0.08)	−0.06 (−0.17, 0.05)	0.004 (−0.07, 0.08)
Non-URM vs Unknown	0.23*** (0.12, 0.34)	0.20*** (0.10, 0.31)	0.26*** (0.12, 0.40)
URM vs Unknown	−0.26*** (−0.37, −0.15)	−0.28*** (−0.39, −0.17)	−0.31*** (−0.46, −0.16)
si:termFall	0.21*** (0.05, 0.37)		
si:GENDERM		0.13 (−0.03, 0.28)	
si:urm_statusNon-URM			−0.12 (−0.34, 0.10)
si:urm_statusURM			0.04 (−0.19, 0.27)
Constant	2.04*** (1.92, 2.16)	2.03*** (1.91, 2.15)	1.98*** (1.86, 2.11)
Observations	3,033	3,033	3,033

Note: *p<0.1; **p<0.05; ***p<0.01

Briefly, SI modifies the effect of semester on GPA, but not gender nor URM status. Further research is needed to better understand what may be going on. An important piece to remember is that for this analysis student level SI data was not available. Thus the "effect" of SI in the analysis shown here may be due to a temporal effect (14-15 vs 16-17).

Using a Logistic model for dfw rate (table not shown here), we see a different impact of SI. Namely gender and URM status significantly differ, not term.