Exploring the Effect of Supplemental Instruction on Equity Gaps: Student

Faith Fatchen, Skip Moses, Rica Rebusit, Joseph Shifman

5/17/2022

Definitions

Two common measurements universities use to identify historically underserved students are underrepresented minority and first generation.

- ▶ **Underrepresented Minority (URM)** - is defined as a U.S. citizen who identifies as Black/African American, Hispanic/Latino, or American Indian. All other Race/Ethnicity categories or Non-U.S. citizens are considered as a Non-Underrepresented Minority (Non-URM).
- ▶ **First generation** - is defined as a student who reported both parents as not receiving a baccalaureate degree. All other students are considered as Not First Generation

Original Source Data (via CSU Chico Instituional Research)

- ▶ Course Detail.csv
  - ▶ n = 43,803
  - ▶ Row contains record for course sections from Fall 2012 to Winter 2022
- ▶ Student Profile Metric.csv
  - ▶ n = 49,107
  - ▶ Row contains one record per student matriculation for students enrolled from Fall 2012 to Spring 2022
- ▶ Student Program.csv
  - ▶ n = 343,781
  - ▶ Row contains records for each semester of each student's attendance who were enrolled from Fall 2012 to Winter 2022

Original Source Datasets (acquired via CSU Cico Instituional Research)

- SLC Appointment.csv
  - $n = 78{,}229$
  - Row contains record for each day that a student went to an SI session, and how many they went to in that day (Fall 2015 - Spring 2022)
- Student Grade.xlsx
  - $n = 864{,}471$
  - Row contains final grade data earned for each course for each student from Fall 2016 to Winter 2022

Data Preparation

Student Level Analysis (n = 19,565)

- ▶ Only student records dated 2016 or later
- ▶ Only first-time freshmen
- ▶ Best attempt at isolating most recent student program record
- ▶ Approximately 35% of students in the programs dataset are not present in the profiles dataset. Therefore, these records have missing data for HS/Transfer GPA, one-year retention, and matriculation/graduation terms
- ▶ Records with missing data for the following attributes were dropped:
    - ▶ High School GPA
    - ▶ Attended Orientation Flag
    - ▶ STEM Major Flag
    - ▶ Full Time / Part Time Code
    - ▶ Academic Program

Data Preparation

Course Level Analysis (n = 5,637)

- Includes class size, average high school GPA, % first-gen in the class, % URM in the class, DWF rate, SI component flag, term year
  - These were calculated using the grades data provided by IR
  - Students with missing records for high school GPA were not included in the calculation
  - Students with missing records for URM and first-gen status were considered as not a member of these groups
- Course records from 2016-2019
- Course sections with less than 10 records of an SI visit during the semester were considered as sections without an SI-component
- Courses represented have a number less than 300 and are not a special number (x89, x99, etc.)
- To remove high calculated DWF outliers, class size $>= 20$

Data Preparation

Coarsened Exact Matching Analysis (n = 19,565)

- ▶ Student profiles, programs, and grades with course level information for courses with SI
- ▶ Only student records dated 2016 to 2019 who DID NOT DROP
- ▶ Best attempt at isolating most recent student program record
- ▶ Approximately 35% of students in the programs dataset are not present in the profiles dataset. Therefore, these records have missing data for HS/Transfer GPA, one-year retention, and matriculation/graduation terms
- ▶ Records with missing data for the following attributes were dropped:
    - ▶ High School GPA
    - ▶ Attended Orientation Flag
    - ▶ STEM Major Flag
    - ▶ Full Time / Part Time Code
    - ▶ Academic Program
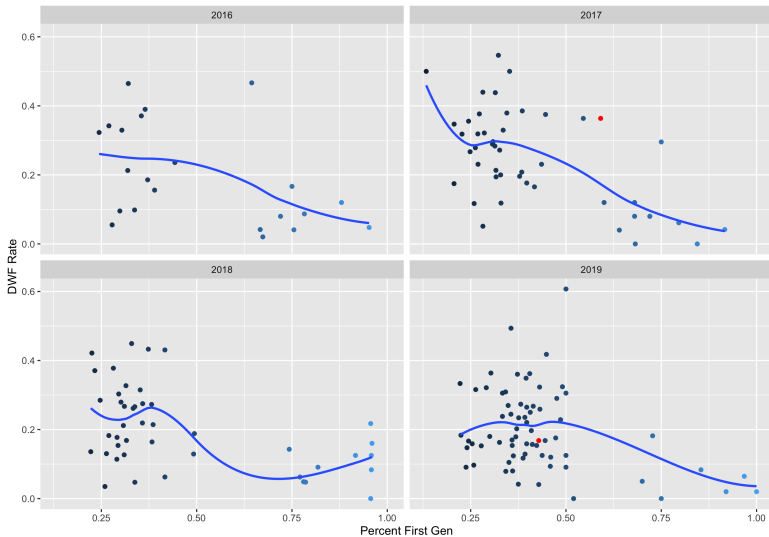
Course Level Details

- ▶ Variables of interest:
  - ▶ Avg HS GPA
  - ▶ First Gen %
  - ▶ URM %
  - ▶ class size
  - ▶ DWF rate
  - ▶ class average
  - ▶ SI component flag
  - ▶ term year
- ▶ Courses from term years 2016-2019
- ▶ Lower Division courses
- ▶ Course sections with class size 20+

URM and DWF Rate
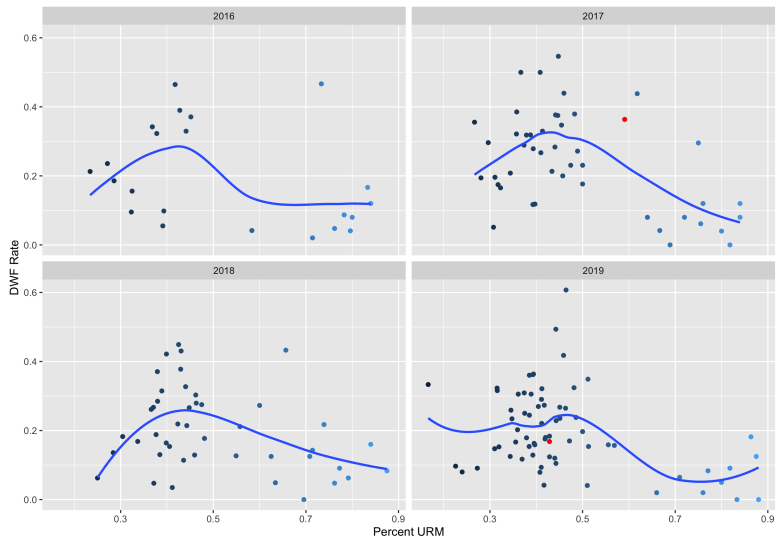
First Generation and DWF Rate

Course Level Modeling

Table 1: Results Multilinear Regression

	Dependent variable:
	dwf.rate
SI.Component.Flag1	0.008 (0.005)
URM	-0.010^{**} (0.005)
First.Gen.Perc	0.009 (0.006)
class.size	-0.0001^{**} (0.00002)
class.average	-0.174^{***} (0.001)
Constant	0.626^{***} (0.004)
Adjusted R^2	0.741
F Statistic	$3{,}231.362^{***}$ (df = 5; 5631)
Note:	*p<0.1; **p<0.05; ***p<0.01

Course Level Modeling

Table 2: Results Multilinear Regression With Interactions

	Dependent variable:
	dwf.rate
SI.Component.Flag1	0.041^{***} (0.014)
URM	-0.012^{**} (0.005)
First.Gen.Perc	0.012^{**} (0.006)
class.size	-0.0001^{***} (0.00002)
class.average	-0.174^{***} (0.001)
SI.Component.Flag1:URM	0.021 (0.044)
SI.Component.Flag1:First.Gen.Perc	-0.096^{***} (0.037)
Constant	0.626^{***} (0.004)
Adjusted R^2	0.742
F Statistic	$2{,}314.955^{***}$ (df = 7; 5629)

Student Level Analysis

- ▶ Measuring student level equity gaps
  - ▶ Students who under represented minorities
  - ▶ Students who are first generation
- ▶ Measuring success
  - ▶ One-year retention
- ▶ First time freshmen matriculated 2016 or later

Student Level Analysis
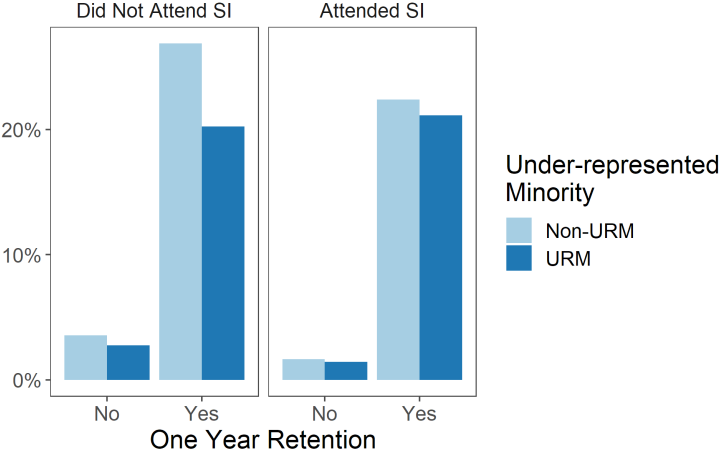
Student Level Analysis

Student Level Analysis: Modeling

- Logistic regression
- Explanatory variables:
  - Student characteristics: URM, First Generation, Gender, Stem Major, number of units taken in the semester, number of units passed in the semester, Cohort term year, SI attendance
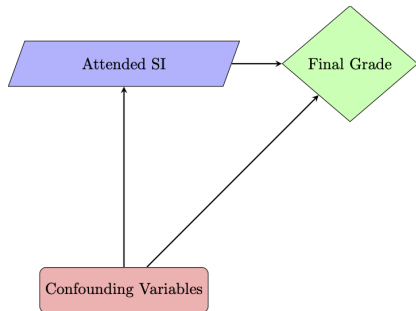 - Course characteristics: Academic level, Course fee existence, GE class

SI and Equity
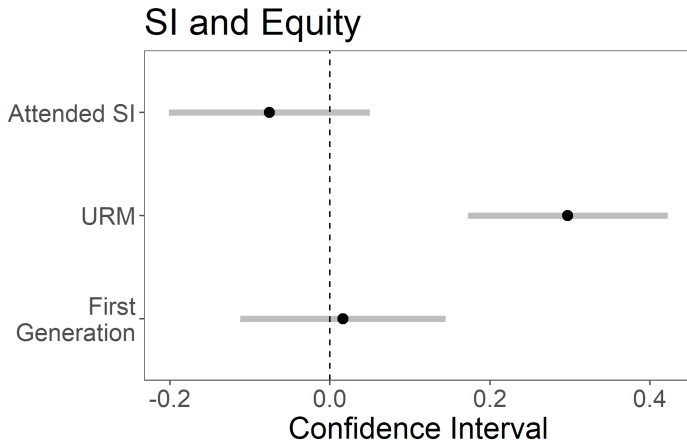
Student Results: Interactions

Interactions SI and Equity
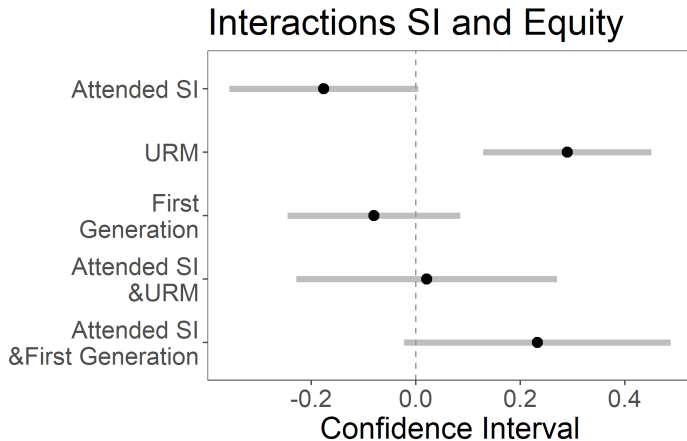
Causal Inference

- We want to measure the impact SI has on student performance. - We need to control for these confounding variables! - We can accomplish this by matching students with similar characteristics.

How do we match?

- ▶ We can't match exactly!
- ▶ We coarsen our covariates (think binning).
- ▶ We can now match exactly on our coarsened data.
- ▶ Lastly, throw out any unmatched observations

Controlling for Confounding with Coarsened Exact Matching

Student 1 ■   ■ Student 6

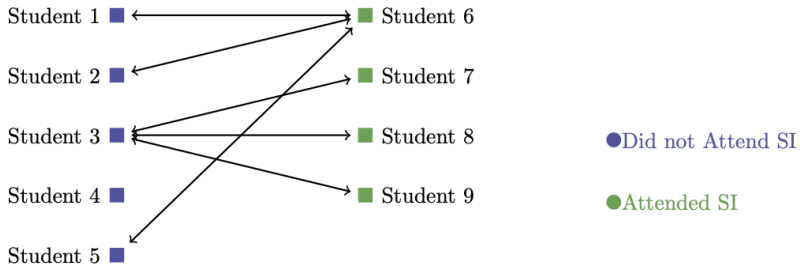Student 2 ■   ■ Student 7

Student 3 ■   ■ Student 8   ● Did not Attend SI
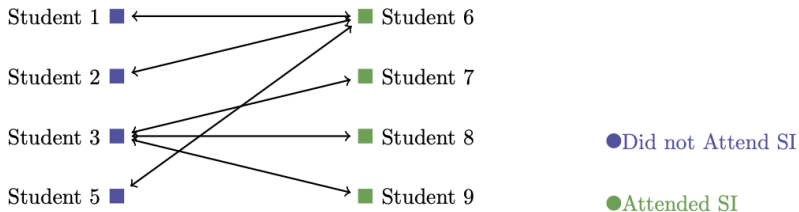
Student 4 ■   ■ Student 9   ● Attended SI

Student 5 ■

Controlling for Confounding with Coarsened Exact Matching

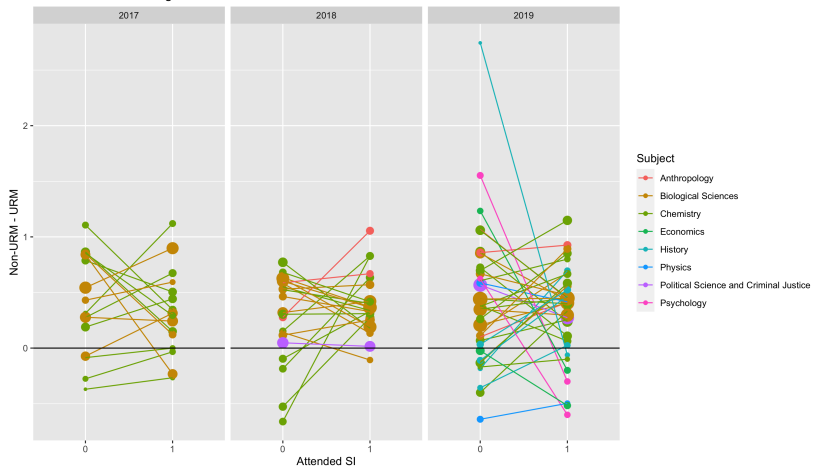# Controlling for Confounding with Coarsened Exact Matching

Covariates for Matching

- Notion: Good students go to SI, so the benefit is inflated.
  - High School GPA: Helps address the notion SI is only seems benefical.
  - Student Orientation Flag:
  - Major Stem Flag
- Demographics
  - Random Course ID
  - Academic Level
  - First Generation Flag
- Equity Gaps
  - URM/NonURM Flag
  - Gender Code

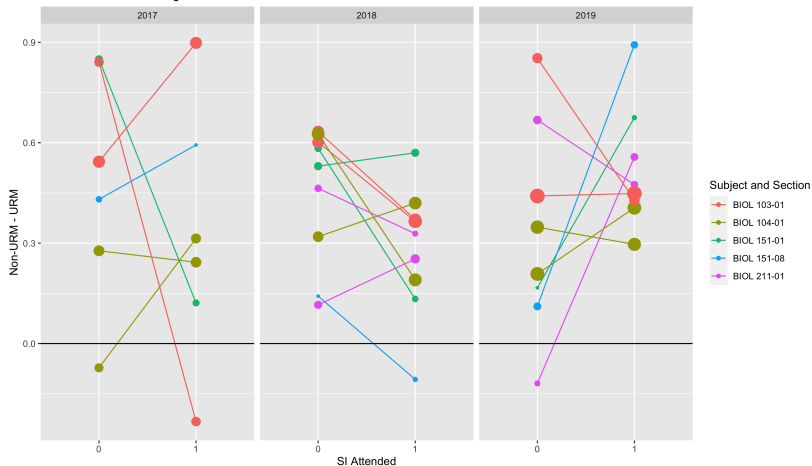# Visualizing the Equity Gap

URM gap among matched students who went to SI
vs those who did not go to SI

Visualizing the Equity Gap

Effect of SI in General

► After creating our matching, we fit a linear model to estimate the effect SI has on a students final grade.
► We find, after controlling for our confounding variables, SI improves students final grade by half of a letter grade.
► This doesn't address the question of whether or not SI reduces equity gaps.

Limitations

- ▶ We have no data on student living situations.
- ▶ The Oroville dam flooding and Camp fire are not being accounted for.
- ▶ We are treating equity gaps separtely.

Next Steps

▶ Meet with an expert in Casual Inference.
▶ Can we extended Causal Inference Techniques to account for equity gaps?
▶ Can we put dollar amount on the benefit of SI by extending the results of the Casual Inference analysis.