

Selective Test-Optional Private University Experimental Protocol

1. The researchers begin with admissions data for students who applied in the 2019-2020 and 2021-2022 cycles and whose application was read twice
 - a. Remove 'special' applicants like recruited athletes and Questbridge applicants, as well as all international students.
2. The researchers randomly select 10 'super-geos', the geographic units that applications are typically broken into for the purpose of reader evaluations, weighting each super-geo by its number of 2020 applicants. New York City (minus Queens) is selected as a pilot super-geo, and an eleventh super-geo is chosen as an 'alternate' in case the admissions office deems one of the selected super-geos to be unusable or particularly unusual. The resulting super-geos are numbered G1 to G10.
3. They then randomly select 60 applications from each super-geo, overweighting applicants from disadvantaged high schools (defined as having a College Board Landscape score of 50 or below) by 50 percent and equivalently reducing weight among applicants who were neither disadvantaged themselves (defined as being either from an underrepresented minority group or being both first-generation and low-income) or attended a disadvantaged high school. Applications are drawn from 2019-2020 applications in groups G1-G8 and from 2021-2022 in G9-G10. Applications that are deemed by the admissions office to be unusable or particularly unusual (e.g. if the applicant would be highly memorable to application readers) will be omitted.
4. They then assign applications to readers as follows:
 - a. For each cluster G#, randomize the students within that cluster and break them into four equally-sized sub-clusters, G#-a to G#-d. Then for each subcluster, define four SAT cases, rounding each a to the nearest 10:
 - i. G#-x-i: Replace SAT with SAT plus randomly-drawn $a \in [-200, -80]$.
 - ii. G#-x-ii: Replace SAT with SAT plus randomly-drawn $a \in [-70, 70]$.
 - iii. G#-x-iii: Replace SAT with SAT plus randomly-drawn $a \in [80, 200]$.
 - iv. G#-x-iv: Omit SAT score.
 - v. For applications with an original score s greater than 1500, we will randomly select three scores in the ranges $[s, 1600]$, $[1400, s]$, and $[1300, 1400]$, labeling these three applications as G#-x-i, -ii, and -iii respectively.
 - vi. For applications with an original score s between 1200 and 1290 (lower than 1200), we will randomly select three scores in the ranges $[1200, s]$, $[s, 1400]$, and $[1400, 1500]$ ($s, [1200, 1300], [1300, 1400]$), labeling these three applications as G#-x-iii, -ii, and -i respectively.
 - vii. In order to report adjusted math and verbal scores separately, we will randomly distribute any adjustment between the two portions (where the shares are drawn from a $[0.25, 0.75]$ uniform distribution for SAT, rounding to the nearest 10, and $[-1, 1]$ for ACT). For instance, consider an applicant with an original SAT score of 1480 (750M 730V); if the new randomly chosen score is 1400 and the random share is 0.5, we will report the score as 1400 (710M 690V). We will disproportionately adjust one score compared to the other if the above procedure would result in an individual portion score greater than 800; in the above example, if the randomly chosen score is 1590 we will report 1590 (800M 790V).

- viii. We will follow the same procedure for applicants that reported an ACT score using the adjustment ranges [-6, -3], [2, 2], and [3, 6]. We will apply this adjustment to each component of the test identically, taking care such that all component scores remain no higher than 36.
- b. Now recombine these pieces by defining four versions of G-# as follows, reordering each back into the original clusters:
- $G\#-1 = G\#-a-i + G\#-b-ii + G\#-c-iii + G\#-d-iv$
 - $G\#-2 = G\#-a-ii + G\#-b-iii + G\#-c-iv + G\#-d-i$
 - $G\#-3 = G\#-a-iii + G\#-b-iv + G\#-c-i + G\#-d-ii$
 - $G\#-4 = G\#-a-iv + G\#-b-i + G\#-c-ii + G\#-d-iii$
- This will result in four versions of each cluster, evenly rotating the applications selected for a large increase, small change, large decrease, or omission of the test score.
- c. Finally, assign packet versions to readers as follows, taking care that no reader is assigned to read any application that they have seen in previous years:

	Test-Optional		Test-Mandatory
	Day 1	Day 2	Day 3
<u>Group 1</u>			
Reader 1	G1-1	G6-2	G3-3
Reader 2	G2-1	G7-2	G4-3
Reader 3	G3-1	G8-2	G5-3
Reader 4	G4-1	G9-2	G1-3
Reader 5	G5-1	G10-2	G2-3
<u>Group 2</u>			
Reader 6	G6-3	G5-4	G2-1
Reader 7	G7-3	G4-4	G1-1
Reader 8	G8-3	G3-4	G5-1
Reader 9	G9-3	G2-4	G4-1
Reader 10	G10-3	G1-4	G3-1
<u>Group 3</u>			
Reader 11	G2-2	G1-3	G5-4
Reader 12	G3-2	G4-3	G2-4
Reader 13	G6-1	G8-4	G1-2
Reader 14	G7-1	G5-3	G3-2
Reader 15	G10-1	G9-4	G4-2
<u>Group 4</u>			
Reader 16	G9-1	G7-4	G5-2
Reader 17	G8-1	G10-4	G2-2
Reader 18	G5-2	G2-3	G1-4
Reader 19	G4-2	G6-4	G3-4
Reader 20	G1-2	G3-3	G4-4

Notes:

- On Day 3, all students with omitted SAT scores should have their SAT replaced with SAT plus randomly-drawn $a \in [x1, x2]$, where $x1$ and $x2$ are chosen such that the resulting interval is the widest interval contained within -200, 200, and the three SAT score adjustments selected in the other three reads of that application.
 - Readers should be reordered so that they are not assigned to read applications that they have previously read.
- 5. The researchers will then provide two datasets back to the university:
 - a. List of applications, packet groupings, and test score manipulations, including:
 - i. The original BannerID
 - ii. New experimental ID (unique to each version of each application). Each application will be read 4-8 times in different conditions.
 - iii. The experimental application packet ID (e.g., G8-3)
 - iv. Randomly assigned test scores
 - b. List of reader and packet assignments
- 6. The university will then create the applications for readers, including:
 - a. Duplicating each of the original applications
 - b. Replace the actual reported test score with the experimentally assigned score
 - c. Assigning the files to the readers in three batches, corresponding to each day.
 - d. The university's admissions office should also scan the files of the original applications to identify applications that are particularly memorable, unusual, or otherwise inappropriate for inclusion in the experiment. The researchers will provide replacement files where necessary.
- 7. Each reader is tasked with evaluating their assigned three sets of applications. They will begin by spending two days evaluating approximately 100 applications, assigning ratings and recommendations (reject, waitlist, committee) to each. They are provided the following information prior to beginning their first day of evaluations:

XXX Admissions has partnered with professors Zachary Bleemer and John Friedman to study XXX's admissions process. We are asking each of XXX's undergraduate admission officers to evaluate three sets of about 50 applications over three days.

The applications that you will be evaluating are based on those of actual XXX applicants from a year or two ago. However, we would like you to evaluate these applications *as if they were real applications in the most recent admissions cycle and you were the second reader of the application*. Just as in the regular admissions process, you will assign each application an academic and non-academic rating. You will also answer the following question for each application: "Do you think this applicant would have been admitted, waitlisted, or rejected this year?"

The purpose of this study is to better understand how XXX's admissions process works in practice and how it has evolved over time. While the Admissions

Office leadership will be able to see your evaluations of each application, your responses will be anonymized before being provided to the researchers.

They will then evaluate their set of applications in Slate.

8. Next, each reader will spend a third day reading approximately 50 evaluations as if XXX returned to a SAT-mandatory admissions policy. In particular, readers will be provided the following information prior to beginning their third day of evaluations:

On this final day of your participation in this study of XXX's admissions process, we'd like you to imagine that XXX has announced a return to its former test-mandated admissions policy. As a result, you'll see that all of the applications you're evaluating today will include standardized test scores. Please evaluate these applications as you would have for the current admissions cycle if XXX had implemented a test-mandated admissions policy.

9. The university will produce the following dataset and provide it to the researchers:
 - a. A dataset with one row for each applicant-reader pair (3,000 total), containing the following fields: experimental application ID, reader ID, time spent on application and any other contextual information collected from Slate, and the reported ratings and recommendation for the applicant.