Why Exclude Test Scores from Admission Criteria? *

Yucheng Liang† Wenzhuo Xu‡

August 20, 2024

Abstract

One major argument in support of test-optional and test-blind college admission policies is that standardized test scores are not a valid measure of students' abilities and are biased against those with fewer resources. This argument goes against standard economic reasoning, which asserts that information, even if invalid or biased, never has negative value. In an experiment, we show that participants responsible for admitting students to an educational opportunity are indeed willing to exclude invalid or biased test scores from admission criteria. This preference is primarily driven by concerns of procedural fairness and the perception that these scores are not very useful. However, the exclusion of these test scores becomes less prevalent after participants gain experience in making admission decisions, allowing them to learn about the scores' usefulness.

*This study is approved by CMU IRB in Protocols 2023_00000333 and 2023_00000337. The RCT registry ID is AEARCTR-0012709. Kate Yixin Huang, Jack Tianrui Lin, and Allison Tribendis provided excellent research assistance. Errors are ours.

†Carnegie Mellon University. Email: ycliang@cmu.edu. Corresponding author.

‡Carnegie Mellon University. Email: wenzhuox@andrew.cmu.edu.

1 Introduction

One of the most intensely debated questions in US college admissions is what information should be included as admission criteria. Recently, a significant shift has occurred concerning the use of standardized test scores such as the SAT and ACT. In 1969, Bowdoin College pioneered an admission policy that allowed applicants to omit their standardized test scores.[1] The adoption of this policy increased gradually in the following five decades and surged during the COVID-19 pandemic, with some colleges deciding to completely stop accepting these scores. The prevalence of test-optional and test-blind admission policies continued even after the pandemic ended. For the 2024 admission cycle, over 2,000 US colleges have adopted these policies.[2] However, the debate around these policies reignited recently as several highly selective colleges reinstated their SAT/ACT requirements, reversing their previous decisions.[3]

The central argument for test-optional and test-blind policies focuses on the validity and bias in standardized test scores, two concepts that have been extensively studied in the education literature (AERA/APA/NCME, 2014; Woo et al., 2023). Critics point out that factors like costly test preparation services can muddle the informativeness of test scores, making them an invalid measure of students' abilities and potential. Moreover, unequal access to test prep and other resources can systematically disadvantage less privileged students, including those from low-income households, making the scores a biased measure (Feder and Bello, 2024). A 2021 survey revealed that 42% of US participants disagreed that standardized tests correctly measure a student's academic knowledge and skills, while 51% believed these tests are biased in favor of affluent students.[4]

Despite these valid concerns about standardized test scores, it remains puzzling why policies that voluntarily give up information have gained widespread adoption and public support. Although imperfect, the SAT and ACT can help predict student success even after controlling for other applicant information (Sackett et al., 2012; Chetty et al., 2023). Given that colleges have

[1] https://fairtest.org/test-optional/
[2] https://fairtest.org/test-optional-list/
[3] https://www.nytimes.com/2024/04/11/us/harvard-test-scores-admissions.html
[4] https://theharrispoll.com/briefs/americans-want-to-end-standardized-tests-for-good/

the freedom to use test scores in a nuanced and context-sensitive manner, why would the general public prefer them to forgo this information?

In this paper, we study information choice in an experiment where participants ("spectators") make admission decisions that could affect the education opportunities of real students. We investigate whether the invalidity and bias of test scores lead spectators to exclude them from the admissions process and explore the rationales behind such exclusions. The experimental method is suitable for this research question because it enables the generation of clean variations in the validity and bias of test scores, which is challenging to achieve in purely observational settings. Additionally, it allows us to observe both information choices and admission decisions within participants. Analyzing the relationship between these two types of decisions can help us better understand the motivations behind information choices.

In the main experiment, each spectator is tasked with selecting students from an introductory data analysis course to participate in an advanced course on the same subject. These students come from different family income backgrounds and have all completed two tests for the introductory course, Test 1 and Test 2. Spectators are aware that students from both higher-income and lower-income backgrounds exhibit similar performance on each test in the absence of any test preparation. However, while there is indeed no test prep for Test 1 for any student, some students may have received test prep for Test 2. When making admission decisions, spectators always observe each student's Test 1 score and family income status. In addition, they may have access to students' Test 2 scores. We ask spectators in an incentive-compatible manner whether they want to have this access under two different scenarios. In one scenario (No Prep), no student receives any test prep for Test 2. The other presented scenario is randomly selected from the following three conditions: Invalid Prep, Biased Prep, and Invalid and Biased Prep. In all three conditions, test preparation may enhance the recipient's Test 2 score by as much as 2 points. What differs across these conditions is the recipients of the prep and whether it improves their data analytic skills. Test prep is deemed invalid if it does not enhance skills and biased if exclusively given to higher-income students instead of everyone.

The results show that the preference to have access to Test 2 scores is strongly influenced by the validity and bias of the test prep situation. Before making any admission decisions, only 11% of spectators prefer not to have access to Test 2 scores if no student received any additional test prep. In contrast, when the test prep is invalid, biased, or both, 35%, 46%, and 56% of spectators prefer to exclude Test 2 scores from their own information sets. The preference to exclude Test 2 scores is mostly strict, and the result is robust to several sample selection criteria.

What motivates spectators to decline access to test scores? From their open-ended explanations for their preferences, we find two main justifications for excluding Test 2 scores. First, the inclusion of test scores that are biased and, to a lesser extent, invalidated by test preparation is deemed unfair. Second, because of the invalidity and bias brought by the test prep, Test 2 scores are considered to be not very useful for admission decisions. These justifications align with a straightforward conceptual framework of procedural fairness, wherein spectators weigh the instrumental value of including the test scores against its moral cost. The importance of fairness concerns is also supported by the result of an additional experiment. In the Performance Prediction treatment, the same information, income status and test scores from the intro course, is used for predicting students' performance in the advanced course as opposed to making admission decisions. Spectators receive bonuses for accurate predictions, yet these decisions do not impact students in any manner. Compared to the main experiment, the fraction of spectators willing to exclude invalid and biased test scores decreases by half, which suggests that much of the exclusion behavior in the main experiment is driven by fairness concerns when the welfare of others is at stake.

After a spectator reports their preference for including or excluding Test 2 scores for two test prep scenarios in the main experiment, one scenario is randomly implemented, wherein the spectator makes 7 rounds of admission decisions. For each decision, the spectator selects 3 students out of a group of 8 to participate in the advanced course. In some decisions, spectators have access only to students' family income status and Test 1 scores, whereas in others, Test 2 scores are also disclosed.

Admission decisions can reveal spectators' social preferences and their decision weights on

the two test scores and family income status. Moreover, by associating admission decisions with information preference within participants, we find evidence consistent with the procedural fairness framework. First, the exclusion of Test 2 scores is associated with lower decision weight on these scores when they are revealed. This correlation underscores the relevance of the scores' instrumental value in determining their demand. Second, the exclusion of test scores is not correlated with measures of outcome-focused social preferences such as meritocracy and preference for lower-income students. This corroborates the non-consequentialist nature of the procedural fairness concerns.

While many spectators initially prefer to exclude invalid and/or biased Test 2 scores, it remains a question whether this preference may change after they gain experience in making admission decisions. To answer this question, we elicit spectators' preferences to include or exclude Test 2 scores a second time after they complete 6 rounds of admissions. Compared to the first elicitation, demand for these scores increases across all scenarios that involve test prep. Moreover, this increase in demand is mostly driven by spectators whose admission decisions rely heavily on Test 2 scores. This implies that the experience effect is likely due to spectators learning about the usefulness of Test 2 scores while making admission decisions. These results resonate with the recent reinstatement of test-required admission policies by several colleges. In their statements, many colleges cite their newly-gained knowledge of SAT and ACT scores' predictive power as a reason for this policy change.[5]

In addition to procedural fairness concerns, we also consider two alternative motivations of excluding Test 2 scores. First, spectators may worry that including Test 2 scores would increase the mental costs of making admission decisions. Second, they may be concerned that having Test 2 scores present may bias their admission decisions. Evidence from our experiment does not support these two motivations. First, very few spectators mention these two rationales in their open-ended justifications. Second, the exclusion of Test 2 scores is not significantly correlated with response time for admission decisions, which is inconsistent with the anticipated decision cost hypothesis.

[5]For example, see Caltech's statement. https://www.caltech.edu/about/news/caltech-restores-standardized-test-requirement-for-undergraduate-admission

Third, when asked to provide advice to other spectators on whether to include or exclude Test 2 scores, most spectators give advice that aligns with their own information preference. This is inconsistent with both alternative hypotheses because, under these motivations, the preference to include or exclude test scores should vary based on who makes the admission decisions. The consistency between information preferences for oneself and advice for others also corroborates the generalizability of our results, demonstrating that our findings apply to both individuals directly involved in admission decision-making and those who are not.

The value of test scores should depend on other applicant information available for admission decisions. For example, SAT and ACT scores may become more valuable for admission if high school grades become less informative due to grade inflation. Conversely, the scores' value may be diminished if colleges are not allowed to adjust the scores based on applicant demographic information. We run two additional experiments to study the effects of other applicant information on the exclusion of test scores, one making Test 1 scores less valid and the other concealing students' family income status. Neither of the additional experiments exhibits an exclusion rate that significantly deviates from that observed in the main experiment. This result suggests that spectators judge the value of test scores in isolation without taking into account the potential influence of other applicant information.

While our incentivized and controlled experiment helps us establish that procedural fairness and usefulness concerns drive spectators' preferences for including or excluding test scores, it remains a question whether these considerations also influence people's support for test-blind and test-optional college admission policies in the real world. To answer this question, we ask spectators in our experiment regarding their support for these policies and their reasons for it. Regardless of their support for these policies, the fairness and usefulness of standardized test scores are indeed the main considerations. Moreover, spectators who use fairness concerns to justify their information preferences in our experiment are also more likely to mention fairness as a factor influencing their policy attitudes. These results demonstrate that the motivations we identify for including or excluding test scores in our experiment are relevant to real-world policy attitudes.

Literature review. This study contributes to the literature exploring the motivations and consequences of test-optional and test-blind admission policies. In a review of empirical evidence, Dynarski et al. (2023) conclude that, despite being motivated by standardized tests' invalidity and bias, test-optional and test-blind policies have limited effects on improving the quality and equity of admission outcomes. For example, Borghesan (2023) finds that test-blind policies lead to a small increase in the enrollment of disadvantaged applicants only at less selective universities. The policies also reduce assortative matching on knowledge, which causes a lower completion rate at elite private colleges.[6] Several papers have proposed alternative motivations, including strategic and general equilibrium factors, for omitting test score requirements.[7] Among these studies, Dessein et al. (2023) is closest to ours. They propose that test-optional and test-blind policies may reduce social pressure when society disapproves of the set of students the college admits. While our research also examines public attitudes, we specifically investigate perspectives on admission *policies*, particularly the inclusion of test scores in admission criteria, rather than focusing on disagreements about the resultant admission *outcomes*.

By studying information preferences for admission decisions, we connect two behavioral economics literatures. One literature studies allocation decisions when recipients' qualifications are uncertain (Cappelen et al., 2022, 2023, 2024; Chakraborty and Henkel, 2024) or when their performances are affected by external factors (Gurdal et al., 2013; Falk et al., 2023; Andre, 2024; Bhattacharya and Mollerstrom, 2022). Although our findings on admission decisions contribute to this body of evidence, our research primarily focuses on information preference, whereas these studies examine decisions under exogenously imposed information structures. Another strand of research investigates information acquisition and belief updating in evaluation decisions, such as hiring (Bartoš et al., 2016; Coffman et al., 2024). Distinct from these studies, our research focuses on admission decisions that impact only the outcomes for others, not the evaluators' own payoffs.

[6] Studying a policy similar to test-optional admissions, Exley et al. (2024) find that, when allowed to do so, female college students are less likely to conceal harmful grades from their transcripts than male students. This gender difference in concealment leads to a larger gender gap in GPA under this policy.

[7] For instance, Garg et al. (2023) posit that dropping test requirements may allow schools to access applicants who are unable or unwilling to take standardized tests. Conlin et al. (2013) argue that test-optional policies might enhance the average *submitted* SAT scores of their enrolled students, potentially boosting their rankings.

This focus allows us to isolate the influence of social preferences on information choice. Moreover, these studies do not address information avoidance which is our focus (see Golman et al. (2017) for a review). Our study is also related to the literature on statistical discrimination in the Phelps (1972) tradition, where groups possessing identical qualifications may receive differential treatment due to varying information structures across these groups. For example, Exley and Nielsen (2024) find that evaluators take workers' self-reported confidence at face value when forming beliefs about their performances, overlooking gender disparities in confidence reporting between men and women with equivalent performance levels. Bohren et al. (2023) find a similar result where evaluators' beliefs about workers' performance do not fully adjust for the group-specific bias in performance signals. Our paper takes a step back and asks whether evaluators want to avoid such disparate information structures. We find that even though spectators do adjust for the invalidity and bias in test scores when making admission decisions, many of them are still willing to give up access to this information.[8] In psychology, Fath et al. (2022, 2023) study evaluators' preference to blind themselves from information such as applicants' race and gender in order to avoid making biased hiring decisions. Crucially, they only consider information that is useless for decisions whereas we focus on test scores which are useful. We also consider bias avoidance as a potential motivation for test score exclusion but find evidence against it.

Our paper extends the literature on procedural fairness which shows that people often value fairness in allocation procedures for non-consequentialist reasons (see Trautmann (2023) for a review). In economics, procedural fairness research has predominantly focused on two types of procedures: the allocation of decision rights (e.g., Bartling et al., 2014) and opportunities (e.g., Bolton et al., 2005). With allocation decisions becoming increasingly data-driven, information choice has become an important procedure in the decision-making process. Our findings indicate that considerations of procedural fairness also drive the exclusion of invalid and biased test scores. In computer science, a large and fast-growing literature on algorithmic fairness studies how to

[8]Another difference between our study and Exley and Nielsen (2024) as well as Bohren et al. (2023) is that our spectators' admission decisions affect only the students but not themselves, whereas the belief reports elicited in their studies affect only the evaluators' payoffs but not the workers'.

select and process information to design fair allocation algorithms. Oftentimes, this entails the exclusion of useful but biased information (see, e.g., Yang and Dobbie, 2020), a practice Rambachan et al. (2020) demonstrate as inconsistent with consequentialist principles. Our findings provide a potential explanation for why information exclusion is a popular approach to achieving algorithmic fairness.

Our paper also makes two contributions to the empirical literature on the demand for useful information (Ambuehl and Li, 2018; Charness et al., 2021; Guan et al., 2023; Liang, 2023; Guan, 2023). First, we demonstrate that fairness concerns can influence information demand, offering a novel explanation for why it may deviate from the instrumental value of information. Second, we show that information demand can change as people gain experience in making decisions with and without information, suggesting that biases in information demand could be mitigated through decision-making experience.

The rest of the paper is organized as follows. Section 2 describes the experimental design. Section 3 lays out a simple conceptual framework. Section 4 presents results on admission decisions with and without Test 2 scores. Section 5 presents the main findings on preferences for including or excluding these scores, while Section 6 examines various rationales behind these preferences. Section 7 reports on attitudes toward test policies in real-world college admissions. Section 8 concludes.

2 Experimental Design

To set up the context where spectators make admission decisions, we recruited students from a US university to take an introductory data analysis course.[9] Upon finishing the course, these students took two tests, referred to as Test 1 and Test 2, covering different course content areas.[10] Each test consisted of five questions with a total possible score of 10. For Test 1, students did

[9]The courses in the experiment are not officially offered by the university. They do not count toward students' GPAs and do not appear on their transcripts.

[10]Test 1 covered numerical data analysis and Test 2 covered textual data analysis.

not receive any additional test preparation beyond the course material. However, for Test 2, we provided some students with one of two types of test prep. The first, "skill-enhancing test prep," offered insights to improve their data analysis skills, and the insights were relevant to one of the test questions. The second type, "non-skill-enhancing test prep," simply provided the answer to a random test question without enhancing analytical skills. Both types of test prep could potentially increase a student's score on Test 2 by up to two points.

We recruit participants ("spectators") from Prolific to complete a survey (see screenshots in Appendix E) in which they select students from the introductory course to enroll in an advanced data analysis course.[11] Each spectator makes admission decisions for 7 groups of students, each time selecting three students out of eight. They know that one of the seven groups consists of real students who took our introductory course while the other six groups are fictitious, and their decisions for the real group have a chance of being implemented. For any group, spectators can also choose an option that says "I cannot decide which 3 students to admit," which will forfeit their chance of determining the admission outcomes. When making admission decisions, spectators have access to each student's score from Test 1 and know whether the student came from a higher-income (self-reported family income $\geq$ \$100,000) or lower-income background ($<$ \$100,000). They may, in addition, know the students' Test 2 scores. This information structure is designed to mimic real-world college admissions, where admission officers can review applicants' high school GPAs along with their demographic information, and they may also have access to their standardized test scores.

Before the admission rounds begin, we ask spectators if they prefer to have access to Test 2 scores for their decision-making. Their responses could be yes, no, or indifferent, along with a justification for their choice. Spectators answer this question twice, each time assuming a different scenario regarding Test 2 preparation. The two scenarios are presented in random order. One scenario is a "**No Prep**" condition where no student received any additional preparation for Test 2. The other scenario is randomized between spectators and includes one of the following conditions:

[11] Prolific is an online platform for distributing surveys commonly used by researchers (Eyal et al., 2021).

- **Invalid Prep**: All students received non-skill-enhancing test prep.
- **Biased Prep**: Only students from higher-income backgrounds received skill-enhancing test prep.
- **Invalid and Biased Prep**: Only students from higher-income backgrounds received non-skill-enhancing test prep.

These scenarios are labeled as such because the type and distribution of test prep directly affect the validity and bias of Test 2 scores. Skill-enhancing test prep simultaneously improves students' Test 2 scores and their data analytic skills, ensuring that the scores remain reflective of their true abilities. In contrast, non-skill-enhancing test prep inflates scores without corresponding skill improvement, thereby undermining the validity of the scores. When test prep is only provided to higher-income students, it creates an additional advantage for them, resulting in Test 2 scores that are more biased in their favor.

Before reporting their information preferences, spectators are informed that, for both tests, students from higher- and lower-income backgrounds performed similarly when there was no test prep. For the scenario involving test prep, spectators also read about its type (whether it is skill-enhancing) and its potential impact on scores (up to 2 points). They understand that one of the two scenarios is the true one that describes the actual test prep situation for the students they will select from, and their reported preferences for this scenario will affect whether Test 2 scores are revealed to them.

After reporting their information preferences, each spectator gets to know the true test prep scenario. For this scenario, we ask spectators to confirm their previously stated information preferences by completing a small real-effort task, which entails typing in a sentence.[12] A confirmation would increase the chance that the observability of Test 2 scores adheres to their stated preferences for the majority of groups. Spectators are free to choose not to complete this real effort task.

[12]The sentence is "I want Test 2 scores to be on the report cards" if the spectator prefers to include the scores and "I do not want Test 2 scores to be on the report cards" if she prefers exclusion.

Unbeknownst to the spectators, the last group is the only one for which the observability of Test 2 scores is affected by their reported information preferences. It is also the only group consisting of real students. Whether Test 2 scores are revealed is fixed for the fictitious groups: they are for Groups 4 to 6 but not for Groups 1 to 3. Spectators always make admission decisions for Groups 1 to 3 before Groups 4 to 6. The orders within the first and second three groups are randomized.

Figure 1 lists the test scores and family income status of the students in the six fictitious groups under the No Prep scenario.[13] The student information of these groups is specifically designed to help us identify spectators' admission objectives and understand their use of information. For example, for the first three groups, choosing higher-scoring students over lower-scoring ones reveals a spectator's preference for meritocracy, while selecting a lower-scoring, lower-income student over a higher-scoring, higher-income student reflects a preference for lower-income students. The scores of Groups 4 to 6 are designed to reveal how spectators weigh Test 2 scores against other information to decide whom to admit. In each of these three groups, the two students in the top row dominate the rest in both tests, whereas the four students in the bottom three rows without the underlines are dominated. Moreover, these dominance relationships hold under any potential impact of test preparation. If a spectator admits the two dominant students and rejects the dominated ones, the only choice remaining is between the two "focal" students with lines under their scores.[14] In Group 4, both focal students come from a lower-income background, each having an edge over the other in one of the two tests. The choice between these students reveals how a spectator weigh Test 2 scores against Test 1 scores for lower-income students. Similarly, the choice between focal students in Group 5 can reflect this tradeoff for high-income students. In Group 6, the two focal students come from different income backgrounds, with the higher-income student having a higher Test 2 score. This design can reveal a spectator's decision weight on Test 2 scores relative to income status.

To make an admission decision for a group, spectators first choose how to order the eight

[13]For the other three test prep scenarios, the student information is modified so that the Test 2 scores for students who received test prep are set to be one point higher.

[14]The lines are not shown to the spectators.

Group 1	
Higher-Income students	Lower-Income students
9, -	9, -
8, -	7, -
5, -	6, -
4, -	4, -

Group 2	
Higher-Income students	Lower-Income students
8, -	6, -
8, -	5, -
7, -	5, -
0, -	5, -

Group 3	
Higher-Income students	Lower-Income students
7, -	7, -
7, -	7, -
7, -	7, -
6, -	7, -

Group 4	
Higher-Income students	Lower-Income students
9, 9	8, 8
6, 5	7, 6
5, 5	6, 8
5, 5	5, 5

Group 5	
Higher-Income students	Lower-Income students
9, 8	9, 9
8, 6	6, 5
7, 8	6, 5
3, 4	6, 5

Group 6	
Higher-Income students	Lower-Income students
7, 9	7, 9
7, 8	7, 7
7, 5	7, 5
6, 5	7, 4

Figure 1: Student information of the six fictitious groups

Notes: The numbers in each cell represent a student's scores in Test 1 and Test 2 under the No Prep scenario. In the other three scenarios, one point is added to Test 2 scores for students who receive test prep under these scenarios. For example, in the Invalid and Biased Prep scenario, the top higher-income student's scores in Group 4 become (9, 10) while the top lower-income student's scores remain at (8, 8). Test 2 scores are disclosed for Groups 4 to 6 but not for Groups 1 to 3.

students on the screen without seeing their information. When Test 2 scores are not observable, spectators can sort students by family income and then by Test 1 scores, or in the reverse order. For groups where Test 2 scores are observable, spectators must also decide whether to sort Test 2 scores before or after Test 1 scores. After making these decisions, the students' information will appear in the chosen order. Spectators are required to select exactly three students for admission unless they opt for the "I cannot decide" option.

After completing the admission decisions for the six fictitious groups, we elicit spectators' information preferences for a second time, which is intended to test whether experience with admission decisions affects information preferences. At this point, spectators are informed that their two reports of information preferences, pre- and post-experience, each have a 50% chance of determining the observability of Test 2 scores for the last group of students. Right after this elicitation, we also ask spectators to advise potential future participants who make admission decisions under the same test prep scenario on whether to request access to Test 2 scores. The advice reflects spectators' information preferences when others decide whom to admit.

The last admission decision is made after the advice elicitation. Finally, we survey participants on their attitudes towards test-blind and test-optional admissions policies and collect demographic information.

Besides this main experiment, we also conduct three additional treatments, each deviating from the main design in one aspect. In the Performance Prediction treatment, instead of making admission decisions, spectators are asked to predict the top 3 performers in the advanced data analysis course for each group of students. Spectators are paid an additional bonus of \$3 if their predictions are accurate, but their decisions do not affect the students in any way. This treatment allows us to study information preferences when the students' payoffs are not at stake. The other two treatments examine the effects of other student information on the demand for Test 2 scores. In the Invalid Test 1 treatment, the Test 1 scores of some randomly selected students are inflated, but spectators do not know who these students are. In the Status Blind treatment, students' family income backgrounds are not revealed to the spectators. In these three treatments, we only elicit

spectators' information preferences under two test prep scenarios: No Prep and Invalid and Biased Prep.

Logistics. We recruited 902 spectators from Prolific on December 18, 2023. Of these, 593 participated in the main treatment, 101 in the Performance Prediction treatment, 107 in the Invalid Test 1 treatment, and 101 in the Status Blind treatment. The participation payment is $5 and the median time spent was 17.5 minutes.

3 Conceptual Framework

Why would spectators choose to exclude freely available test score information from admission decisions? In this section, we start from a standard consequentialist model and derive a negative result: even if she cares about fair outcomes, a standard consequentialist spectator would never exclude any information. Then, we consider several deviations from this model which could lead to the exclusion of test scores and derive their implications.

3.1 Set-up

Let θ be the state of the world relevant to admission, such as students' income, abilities, and potential. The admission decision is denoted by $a \in A$, representing whether each student is admitted or rejected. Prior to the decision, the spectator decides whether to have access to students' Test 2 scores $t \in T$, knowing that she will have access to some other applicant information denoted by $s \in S$, which in the main experiment includes Test 1 scores and family income categories. The joint distribution of θ, t, and s is denoted by p.

The spectator's preference for admission outcomes is described by the value function $v(a, \theta)$. Crucially, the inclusion or exclusion of t does not directly enter the value function. Nevertheless, this function is general enough to accommodate social preferences such as meritocracy, preference for lower-income students, preference for diversity, etc.

3.2 Standard consequentialism

A standard consequentialist makes admission decision $a(\emptyset, s)$ to maximize the expected value of admission outcomes when she only has access to s:

$$a(\emptyset, s) = \arg\max_{a} E_{\theta}[v(a, \theta)|s]. \tag{1}$$

Under this information structure, her ex-ante expected value is

$$\begin{aligned} V(\emptyset, S) &= E_s[E_{\theta}[v(a(\emptyset, s), \theta)|s]] \\ &= E_{t,s}[E_{\theta}[v(a(\emptyset, s), \theta)|t, s]]. \end{aligned} \tag{2}$$

When the spectator also knows students' Test 2 scores t, her optimal admission decision becomes

$$a(t, s) = \arg\max_{a} E_{\theta}[v(a, \theta)|t, s] \tag{3}$$

and the ex-ante expected value is

$$V(T, S) = E_{t,s}[E_{\theta}[v(a(t, s), \theta)|t, s]]. \tag{4}$$

Therefore, the value of including Test 2 scores for a standard consequentialist is the difference between expressions (4) and (2):

$$\begin{aligned} V^{SC}(T) &= V(T, S) - V(\emptyset, S) \\ &= E_{t,s}[E_{\theta}[v(a(t, s), \theta) - v(a(\emptyset, s), \theta)|t, s]]. \end{aligned} \tag{5}$$

By the definition of $a(t, s)$ in equation (3), $E_{\theta}[v(a(t, s), \theta) - v(a(\emptyset, s), \theta)|t, s] \geq 0$ for any t and s. Hence, expression (5) is always non-negative, implying that the spectator never strictly prefers to exclude Test 2 scores. Intuitive, this is because standard consequentialists always have the free

option to ignore Test 2 scores when they do not want to use them; when they do use the scores, it must be optimal to do so. As a result, including Test 2 scores as an admission criterion has no cost.

3.3 Procedural fairness

For allocation decisions, people may care about the fairness of the decision procedure independent of the resulting outcome. For spectators with such concerns, including invalid or biased test scores as part of the admission criteria may taint the fairness of the whole procedure. We model procedural fairness concerns as a reduced-form moral cost $C(T)$ of including the test scores T. The value of including Test 2 scores then becomes

$$V^{PF}(T) = E_{t,s}[E_\theta[v(a(t,s),\theta) - v(a(\emptyset,s),\theta)|t,s]] - C(T). \tag{6}$$

With procedural fairness concerns, whether to include or exclude the test scores depends on the tradeoff between the expected usefulness of the scores (the first term) and the moral cost of including them (the second term). Assuming that the expected usefulness is positively correlated with the actual use of the scores in admission decisions, this tradeoff leads to a testable comparative statics prediction: the preference to exclude Test 2 scores is negatively correlated with the use of these scores in admission decisions when they are available.

3.4 Anticipated decision costs

Another potential reason for spectators to exclude Test 2 scores is that they may anticipate admission decisions with Test 2 scores available to be more costly than those without. The cost could stem from the mental effort required to incorporate additional information, and it can be especially high when spectators have to extract information content amid potential noise and biases. In real-world college admissions, the cost could also come from the increased social pressure if the test scores are also observed by third parties (Dessein et al., 2023). To model a spectator who anticipates decision costs, we allow the cost c to depend on the admission decision a, the state θ, and the

information t and s. The spectator chooses $a(t,s)$ to maximize $E_\theta[v(a,\theta) - c(a,\theta,t,s)|t,s]$, and the value of including Test 2 scores is

$$V^{DC}(T) = E_{t,s}[E_\theta[v(a(t,s),\theta) - v(a(\emptyset,s),\theta) - (c(a(t,s),\theta,t,s) - c(a(\emptyset,s),\theta,\emptyset,s))\,|t,s]]. \quad (7)$$

Because of the additional cost terms in equation (7), the value of including Test 2 scores could be negative. Assuming that anticipated decision costs are positively correlated with response time when spectators make admission decisions (Fudenberg et al., 2018; Halevy et al., 2023), one implication of this model is that spectators who prefer to exclude Test 2 scores also spend more time on admission decisions when Test 2 scores are available. Another implication is that spectators would be less willing to exclude the test scores if the admission decisions are made by other people because they no longer bear the decision costs by themselves.

3.5 Anticipated decision mistakes

One last explanation for test score exclusion we consider is related to anticipated decision mistakes. If a spectator is concerned that her admission decision may be biased when she observes Test 2 scores, she may be willing to blind herself from them. Formally, the spectator anticipates that she will choose $\hat{a}(t,s)$ when Test 2 scores are observed, which differs from her optimal choice $a(t,s)$. Therefore, the value of including these scores,

$$V^{DM}(T) = E_{t,s}[E_\theta[v(\hat{a}(t,s),\theta) - v(\hat{a}(\emptyset,s),\theta)|t,s]], \quad (8)$$

could be negative.

This explanation for test score exclusion yields a testable implication when admission decisions are made by others, which is opposite to the prediction of the decision cost explanation. If spectators believe that others' admission decisions are more likely to be misaligned with their own standards, they should be more inclined to exclude Test 2 scores when these decisions are made by others.

4 Results on Admission Decisions

We begin by analyzing the aggregate patterns of spectators' admission decisions for the six fictitious groups. These decisions inform us of the spectators' admission objectives and use of information, which are pertinent to our later analysis of their information choices.

4.1 Admission decisions without Test 2 scores

For the first three groups of students whose Test 2 scores are not revealed, spectators' admission decisions are mostly consistent with meritocracy, with some revealing an additional preference for admitting lower-income students. For Groups 1 and 2, 76% and 62% of decisions are purely meritocratic, admitting three students with the highest Test 1 scores regardless of their income status. In addition, 15% and 25% decisions can be explained by a mix of meritocracy and low-income preference. These decisions are meritocratic within each income group, but may choose lower-income students over their higher-performing, higher-income peers. The preference for lower-income students is rarely absolute, as only 2% and 1% of decisions exclusively admit these students. Together, meritocracy and low-income preference account for around 90% of decisions for Groups 1 and 2. For Group 3 where three higher-income and four lower-income students are tied for the top Test 1 scores, 24% of spectators state that they could not make a decision, which implies that decision costs are substantial when students are not differentiable through test scores.[15] Almost everyone who does make a decision chooses three of the top performers.

4.2 Admission decisions with Test 2 scores

For Groups 4, 5 and 6 where scores of both tests are revealed, admission decisions can reveal how much weight spectators put on Test 2 scores relative to Test 1 scores and income status. The results show that spectators put lower weight on Test 2 scores when the scores are affected by non-skill-enhancing test prep. In addition, the weight on higher-income students' Test 2 scores is

[15] Decision avoidance is very rare for Groups 1 and 2, accounting for only 1% and 2% of the decisions, respectively.

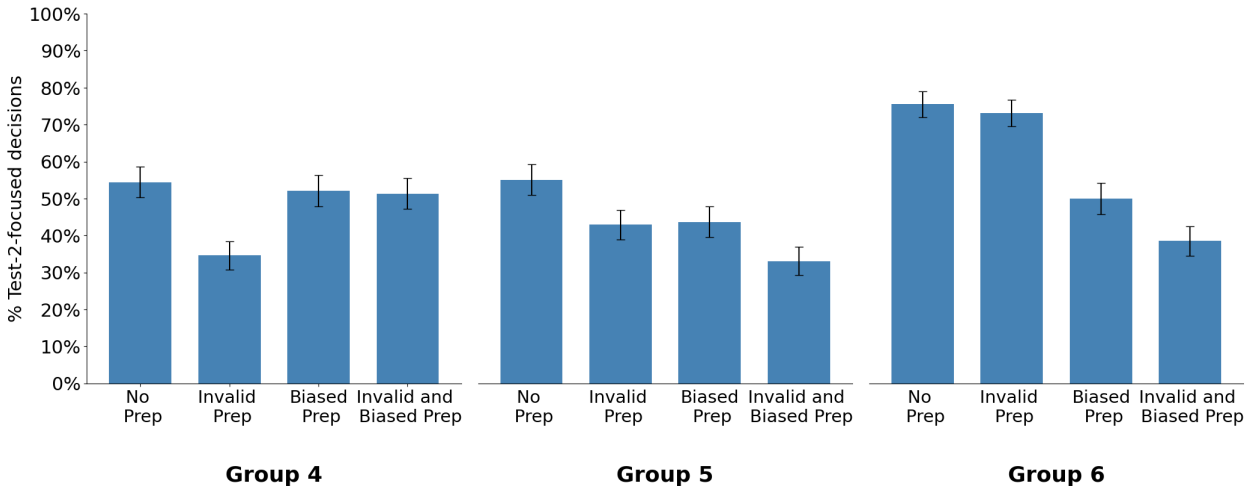

Figure 2: Proportions of Test-2-focused admission decisions for Groups 4 to 6

Notes: This figure shows the proportions of Test-2-focused admission decisions for each group under each test prep scenario. A Test-2-focused decision occurs when the focal student with a higher Test 2 score is admitted and the other focal student is rejected. Error bars represent standard errors.

sometimes lower when they are affected by test prep that is unavailable to lower-income students.

Specifically, as explained in Section 2, each group includes one higher-income and one lower-income student who dominates all other students in both tests. There are also four students in each group whose scores are dominated by the rest. These dominance relationships hold under any potential impact of test preparation. As a result of this design, the vast majority of spectators (91% for Group 4, 89% for Group 5, and 83% for Group 6) admit the two dominant students and reject the dominated, and the third admitted student must come from the remaining two focal students.

In Group 4, both focal students come from a lower-income background, each having an edge over the other in one of the two tests. As the left panel of Figure 2 shows, the proportion of Test-2-focused decisions (admitting the focal student with a higher Test 2 score and rejecting the other) is roughly the same in Scenarios No Prep, Invalid Prep, and Invalid and Biased Prep, but smaller in the Invalid Prep Scenario. This result is consistent with Bayesian updating—lower-income students' Test 2 scores are equally informative in all scenarios except the last one where the scores are noised up by the non-skill-enhancing test prep. Bayes' rule hence dictates that spectators in this scenario should down-weight Test 2 scores.

In Group 5, again, each of the two focal students has a higher score on one test, but they

both come from a higher-income background. Similar to Group 4 decisions, there are fewer Test-2-focused decisions when the scores of this test are affected by non-skill-enhancing test prep (No Prep − Invalid Prep $= 12\%$, $p = 0.02$). In addition, in scenarios where these higher-income students receive test prep that is not available to their lower-income peers, the decision weight on Test 2 scores is further lower (Invalid Prep − Invalid and Biased Prep $= 10\%$, $p = 0.04$). The latter effect is not explainable by Bayesian updating because whether lower-income students receive the test prep or not does not affect the informativeness of the two higher-income students' scores. Hence, this effect implies that the effect biased test prep has on spectators' decisions goes beyond its effect on the informativeness of the scores.

In Group 6, the two focal students have the same Test 1 score. One is a higher-income student with a higher Test 2 score, while the other is a lower-income student. Spectators are less likely to select the higher-scoring, higher-income student over his lower-income peer if he receives exclusive test prep (No Prep − Biased Prep $= 26\%$, $p < 0.001$), especially if the test prep does not enhance skills (Biased Prep − Invalid and Biased Prep $= 11\%$, $p = 0.02$).

5 Information Preferences under Different Test Prep Scenarios

Do spectators prefer that Test 2 scores be revealed when they make admission decisions? Figure 3 shows the distributions of information preferences under different test prep scenarios before spectators have any experience making admission decisions. 11% of participants prefer to exclude Test 2 scores from the admission process in the No Prep scenario, whereas the number increases to 35%, 46%, and 56% when the test prep is invalid, biased, and both, respectively. The choice to exclude Test 2 scores mostly reflects strict preferences: 92% of spectators who make this choice, when prompted, complete the real effort task to confirm their exclusion preference (Figure A1 shows the distributions of confirmed information preferences). Moreover, this finding is robust to the order in which the scenarios are encountered (see Figure A2). Regression analysis shows that the effects persist after controlling for demographics or spectator fixed effect (see Table A1).

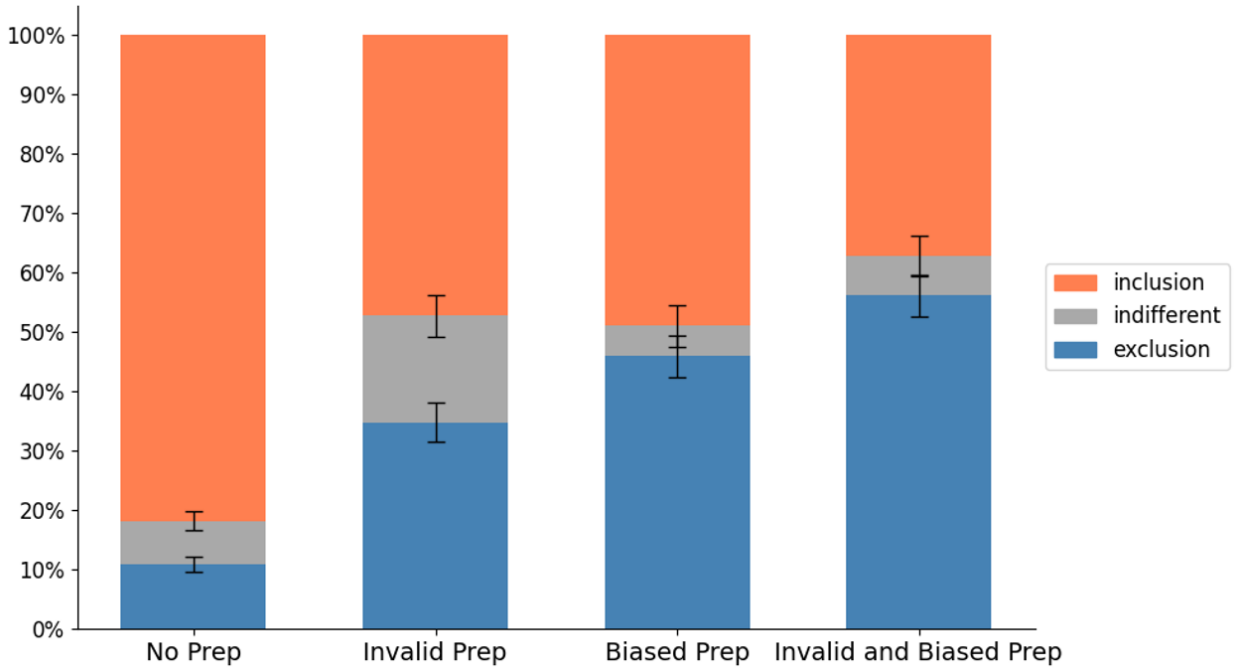

Figure 3: Information preferences across test prep scenarios (before admission experience)

Notes: This figure shows the proportions of spectators who prefer to include, exclude, or are indifferent about the inclusion of Test 2 scores under each test prep scenario. Error bars represent the standard errors of the proportions of inclusion and exclusion preferences.

Taken together, these results indicate that the invalidity and bias of test scores make people more willing to exclude it from the admission process.

6 Rationales for Excluding Test 2 Scores

6.1 Self-reported rationales

To understand the rationales behind the information preferences, we ask spectators to provide justifications in an open-ended question and summarize the answers using GPT-3.5, a large language model (see Appendix B for details of the summarizing procedure). The exclusion of Test 2 scores predominantly revolves around two considerations – the fairness and usefulness of the test scores.[16] We then hand-code each answer as to whether it mentions these two considerations.

[16]The main rationale for including Test 2 scores, on the other hand, focuses on the usefulness of the scores.

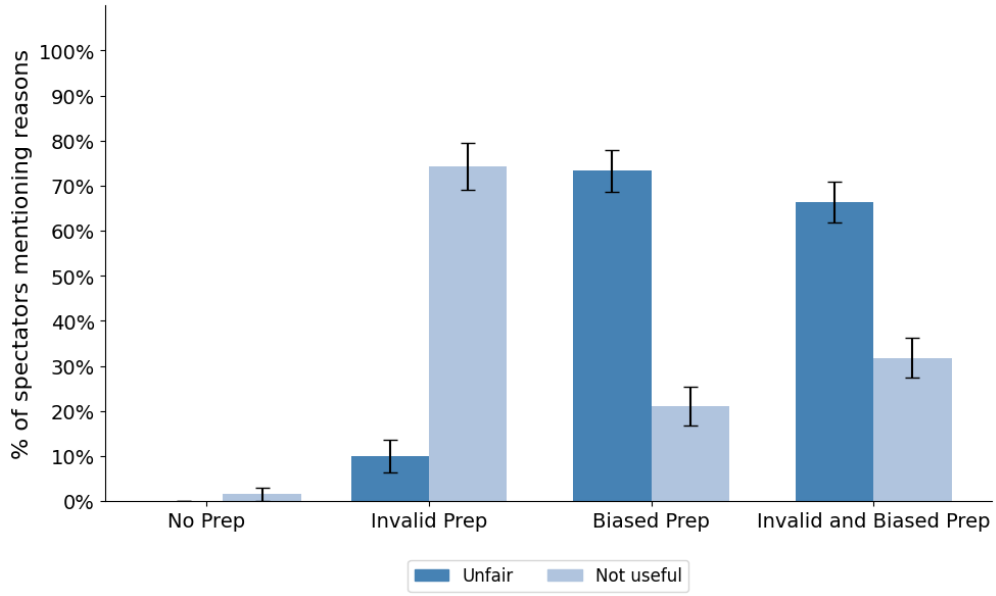

Figure 4: Self-reported reasons for excluding Test 2 scores

Notes: This figure shows, for each test prep scenario, the proportions of justifications for excluding Test 2 scores that mention the scores being "unfair" or "not useful." Error bars represent the standard errors of the proportions.

Figure 4 shows the prevalence of the two rationales in each scenario. In the two scenarios where only higher-income students receive test prep, 73% (Biased) and 66% (Invalid and Biased) of spectators who prefer to exclude Test 2 scores mention that including the scores would be unfair. This number drops to 10% in the Invalid Prep scenario, and no spectators mention unfairness in the No Prep scenario. In the three scenarios with test prep, 74% (Invalid), 21% (Biased) and 32% (Invalid and Biased) of spectators who prefer to exclude Test 2 scores mention that the scores would not be useful for their admission decisions. For the No Prep scenario, the number drops to 2%. In addition to "unfair" and "not useful," we also look for answers that mention decision costs and decision mistakes as justifications for excluding Test 2 scores. Across all spectators and scenarios, only a handful of answers mention these considerations.

The result on spectators' self-reported justifications for test score exclusion strongly supports the procedural fairness model outlined in Section 3. Under this model, spectators face a trade-off between the moral cost of including unfair test scores and their usefulness for making more informed admission decisions. Therefore, fairness and usefulness emerge as the two main con-

siderations that drive information preferences. In the rest of this section, we provide additional behavioral evidence for these two rationales.

6.2 Fairness Concerns and the Exclusion of Test 2 Scores

To provide behavioral evidence on fairness concerns as a driver of the exclusion of Test 2 scores, we conduct a diagnostic treatment (Performance Prediction) where spectators' decisions do not affect the students in any way, making other-regarding preferences irrelevant. Specifically, spectators are asked to predict the top 3 performers in the advanced data analysis course for each group of students. The predictions are incentivized for accuracy with a potential bonus of \$3. Same as in the main treatment, each group consists of 4 higher-income and 4 lower-income students who have taken two tests for the introductory course, and spectators always observe students' income backgrounds and Test 1 scores. We elicit spectators' preferences for including or excluding Test 2 scores under two test prep scenarios: No Prep and Invalid and Biased Prep.

In this treatment, only 29% of spectators prefer to exclude Test 2 scores in the Invalid and Biased Prep Scenario, and among them only 90% confirm their preferences by completing the real-effort task. This is significantly less prevalent than the 56% of spectators who prefer exclusion in the main treatment under the same scenario ($p < 0.001$). The finding that spectators become much less likely to exclude invalid and biased test scores when their decisions do not affect the students implies that fairness concerns are an important driver for the exclusion preference.

One may wonder whether the fairness concerns that matter for test score exclusion are about the admission procedure or the admission outcomes. Theoretically, as we show in Section 3, concerns about outcome fairness alone cannot generate a strict preference for excluding test scores. This is because when test scores are revealed, spectators could always choose not to use them, which guarantees a lower bound of outcome fairness that is the same as what could be achieved without the test scores. Empirically, Table 1 shows that preferences to include or exclude Test 2 scores for any scenario with test prep are uncorrelated with two measures of outcome-based social preferences, the number of meritocratic decisions and the number of admitted lower-income

	Information Preference (include: 1; indifferent: 0; exclude: -1)			
	No Prep	Invalid Prep	Biased Prep	Invalid and Biased Prep
# of meritocratic decisions	0.06	0.001	-0.08	-0.19
	(0.04)	(0.10)	(0.10)	(0.10)
# of lower-income admissions	0.02	0.05	-0.04	-0.06
	(0.02)	(0.05)	(0.05)	(0.05)
R^2	0.004	0.01	0.003	0.02
Observations	593	201	196	196

Table 1: Information preferences (before admission experience) and outcome-based social preferences

Notes: This table shows OLS estimates of regressing information preferences before admission experience on outcome-based social preferences revealed from admission decisions for Groups 1 to 3. The dependent variables are information preferences (include: 1; indifferent: 0; exclude: -1) under each scenario. Numbers in brackets are standard errors. $^{***}: p < 0.001$; $^{**}: p < 0.01$; $^{*}: p < 0.05$

students for Groups 1 to 3. This result indicates that the fairness consideration that drives the exclusion of Test 2 scores is likely concerned with the admission procedures, not its outcomes.

6.3 Usefulness and the Exclusion of Test 2 Scores

To provide behavioral evidence on the association between test score exclusion and the use of test scores when they are available, we compare the proportions of Test-2-focused decisions between spectators who prefer to exclude Test 2 scores and those who do not. As Figure 5 shows, across Groups 4 to 6 and across the four test prep scenarios, spectators who choose to exclude Test 2 scores use them less when they are available. Again, this result is consistent with the procedural fairness model.

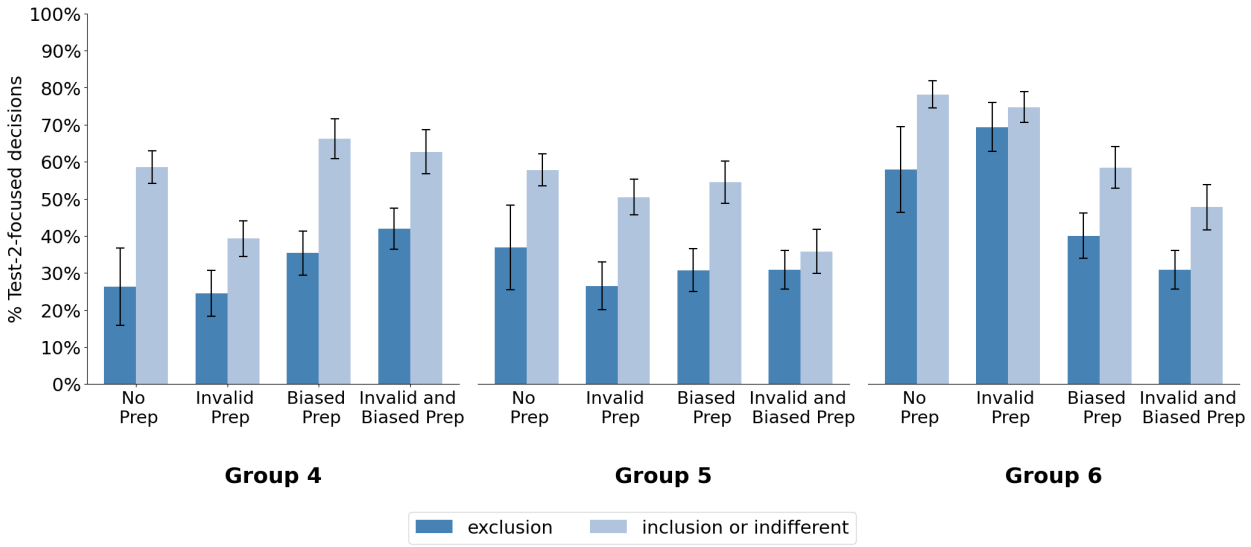

Figure 5: Proportions of Test-2-focused admission decisions by information preferences

Notes: This figure shows the proportions of Test-2-focused admission decisions for each group under each test prep scenario, separately for spectators who prefer to exclude Test 2 scores and those who do not. A Test-2-focused decision occurs when the focal student with a higher Test 2 score is admitted and the other focal student is rejected.

6.4 Admission Experience and the Exclusion of Test 2 Scores

Because we elicit spectators' information preferences a second time after six rounds of admission decisions, we can study whether information preferences change after spectators gain experience in making admission decisions. As shown in Figure 6, after six admission rounds, demand for Test 2 scores increases across all three scenarios that involve test preparation (see also Figure A3 for the transition matrix of information preferences before and after admission experience for each scenario). This indicates that decision-making experience can reduce the preference to exclude test score information. Moreover, Table 2 shows that this increase in demand is more pronounced for spectators who make many Test-2-focused decisions for Groups 4 to 6. This implies that the experience effect is likely due to spectators learning about the usefulness of Test 2 scores during their decision-making process.

The experience effect in our experiment resonates with the recent trend among several selective colleges that have reinstated their standardized testing requirements for admissions. In justifying this reversal, most of these institutions have highlighted the value of SAT/ACT scores. For in-

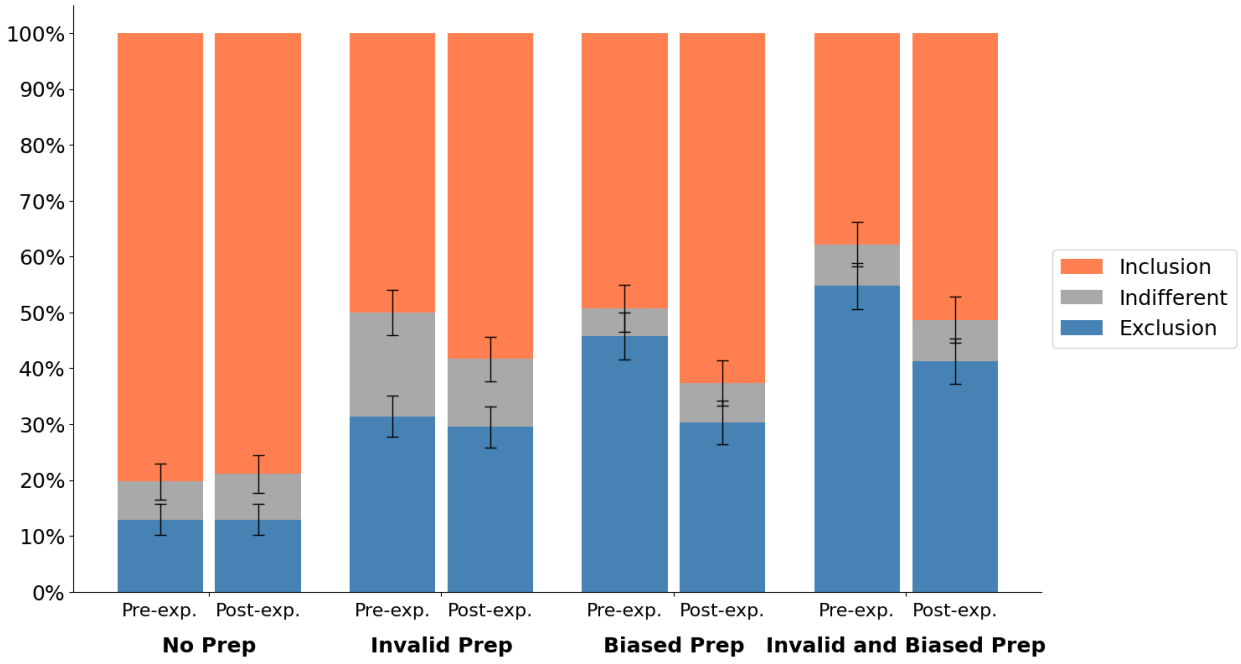

Figure 6: Information preferences across test prep scenarios (before and after admission experience)

Notes: This figure shows the proportions of spectators who prefer to include, exclude, or are indifferent about the inclusion of Test 2 scores under the true test prep scenario. For each scenario, the bar on the left represents information preferences before admission experience, and the bar on the right represents information preferences after admission experience. Error bars represent the standard errors of the proportions of inclusion and exclusion preferences.

	Post-experience Information Preference			
	No Prep	Invalid Prep	Biased Prep	Invalid and Biased Prep
# of Test-2-focused decisions	0.14**	0.17**	0.17**	0.20***
	(0.05)	(0.06)	(0.05)	(0.06)
Pre-exp. information preference	0.47***	0.52***	0.56***	0.58***
	(0.07)	(0.07)	(0.06)	(0.06)
R^2	0.31	0.36	0.48	0.43
Observations	147	156	142	147

Table 2: The use of Test 2 scores and post-experience information preferences

Notes: This table shows the OLS estimates of regressing information preferences after admission experience on the numbers of Test-2-focused decisions for Groups 4 to 6, controlling for information preferences before admission experience. Information preferences are coded as follows: include: 1; indifferent: 0; exclude: -1. A Test-2-focused decision occurs when the focal student with a higher Test 2 score is admitted and the other focal student is rejected. Numbers in brackets are standard errors. $^{***}: p < 0.001$; $^{**}: p < 0.01$; $^{*}: p < 0.05$

stance, Caltech stated that "standardized testing provides admissions officers and faculty reviewers useful information about academic preparedness."[17] This marks a significant shift from a 2022 announcement that the test-blind policy would continue until at least the 2025 admission cycle, indicating that new insights have been gained since then.[18]

6.5 Anticipated Decision Costs and Decision Mistakes

In this section, we consider two alternative motivations for excluding Test 2 scores: spectators may worry that including Test 2 scores would 1) increase the cost of making admission decisions or 2) increase the likelihood of decision mistakes. As we discuss in Section 6, very few spectators mention these two motivations in their open-ended justifications for their decisions. We will now

[17] https://www.caltech.edu/about/news/caltech-restores-standardized-test-requirement-for-undergraduate-admission

[18] Similarly, Harvard recently ended its test-optional policy, at least two years earlier than previously expected, citing the usefulness of test scores in identifying talent.

present more behavioral evidence that is inconsistent with these motivations.

6.5.1 Decision time

We focus on the time a spectator takes to submit her admission decision as a measure of decision costs.[19] As Table A2 shows, the average decision time for an admission decision with Test 2 scores available is not significantly correlated with information preference either before or after the six rounds of admission. The correlations are also mostly small and insignificant if we consider the difference in decision time between decisions with and without Test 2 scores. These results indicate that decision costs are not a main driver of the exclusion of Test 2 scores.

6.5.2 Advice

We ask spectators to advise other spectators on whether to include Test 2 scores and explain their advice. The comparison between advice and spectators' own information preferences can test the two alternative motivations. If a spectator prefers to exclude Test 2 scores because she wants to lower the costs of making admission decisions, she should be less likely to advise others to exclude the scores as the costs are no longer borne by herself. Conversely, if she is concerned about her admission decisions being biased by the Test 2 scores, then she should be more willing to advise others to exclude the scores assuming that others' admission decisions are less likely to be aligned with her ideals.

Figure 7 cross-tabulates the advice against spectators' own information preferences elicited right before the advice. First and foremost, few spectators give advice opposite to their own preferences. Among the 169 spectators who prefer to exclude Test 2 scores for themselves, only 13 advise others to include the scores. Moreover, as is shown in Figure A4, nearly half (6) of these reversals come from the No Prep scenario, suggesting that decision costs are not a main reason for excluding invalid or biased Test 2 scores. For spectators who prefer inclusion for themselves,

[19] An alternative measure is decision avoidance, i.e., selecting "I cannot decide which 3 students to admit." Decision avoidance is very rare when Test 2 scores are available (1% for Group 4, 1% for Group 5, and 5% for Group 6), so there isn't sufficient variation in this measure to test the effect of decision costs on information preference.

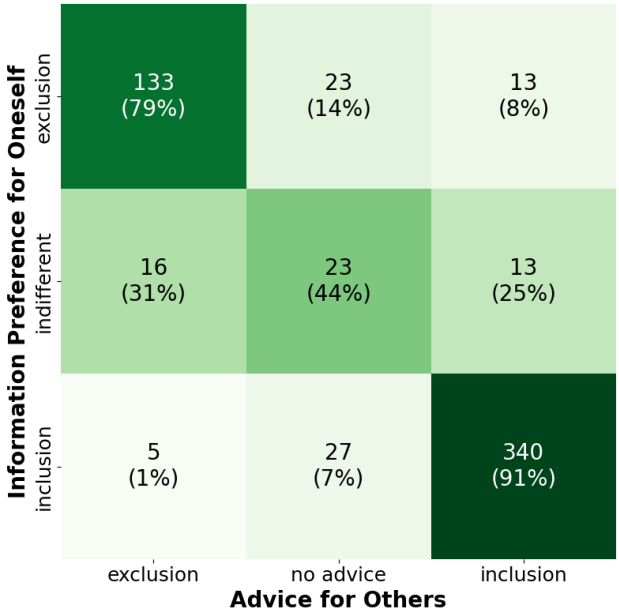

Figure 7: Transition matrix for information preferences for oneself and advice for others

Notes: This figure shows the numbers of spectators in cells defined by their information preferences for themselves (after admission experience) and advice to others. Numbers in brackets represent proportions relative to the number of spectators with the same information preference.

advice in the opposite direction is even rarer, suggesting that anticipated decision mistakes are unlikely to be a pivotal factor for information preferences. 50 spectators who are either in favor of or against including Test 2 scores themselves decide not to provide any advice to others. A significant proportion (26%) of this behavior is justified by respect for others' agency. Among those who express indifference for themselves, 25% advise for inclusion and 31% advise against it. Taken together, the evidence suggests that anticipated decision costs and decision mistakes are not key drivers of the exclusion of Test 2 scores.

The results on advice are also important for the realism of our study as many people who express opinions on college admission policies are likely not making admission decisions themselves. The fact that spectators' own information preferences and their advice are highly consistent supports our findings' external validity on that aspect.

6.6 Other Student Information

An important factor that affects the value of standardized test scores for admission is the availability and quality of other applicant information such as high school GPAs, application essays, and demographic and socioeconomic information. In fact, both proponents and critics of test-optional and test-blind policies have cited the information environment as part of their arguments. For example, the two sides have debated about the predictive accuracy of high school grades for college success under the same premise that test-optional and test-blind policies make more sense when these grades are more informative. [20] Also, after the Supreme Court's decision that bans the use of race in college admissions, some commentators predict that more colleges will stop requiring standardized test scores because the value of these scores is diminished when colleges are no longer able to adjust for the racial gap.[21]

To investigate the effects of other applicant information on preferences to include or exclude Test 2 scores, we run two additional treatments, both focusing on the Invalid and Biased Prep scenario. The Invalid Test 1 treatment (N=107) differs from the main treatment in that Test 1 scores are randomly inflated but spectators don't know who the beneficiaries are. By comparing the preferences to include or exclude Test 2 scores between this treatment and the main treatment, we can test whether the demand for one test score depends on the validity of other admission criteria. In the Status Blind treatment (N=101), spectators are not informed about students' family income background. This treatment can test if demand for test scores decreases when spectators cannot use other applicant information to adjust for the scores' bias.

Figure 8 shows that the demand for Test 2 scores in these two treatments is the same as the main treatment. This result indicates that although the value of test scores should depend on other

[20] For example, the CEO of ACT said that "the score is just one measure of student success—in the face of systemic, persistent grade inflation, it's an increasingly critical one."(https://www.insidehighered.com/news/admissions/traditional-age/2024/01/17/reigniting-standardized-testing-debate) On the one hand, FairTest.org justified test-optional and test-blind policies by citing a study that the grades are quite informative predictors especially within demographic categories (https://fairtest.org/wp-content/uploads/2024/03/TestOptionalReportFinal.pdf).

[21] https://www.forbes.com/sites/vinaybhaskara/2023/07/05/how-the-end-of-affirmative-action-will-impact-college-admissions/?sh=bc5dfea32db7. See also Chan and Eyster (2003).

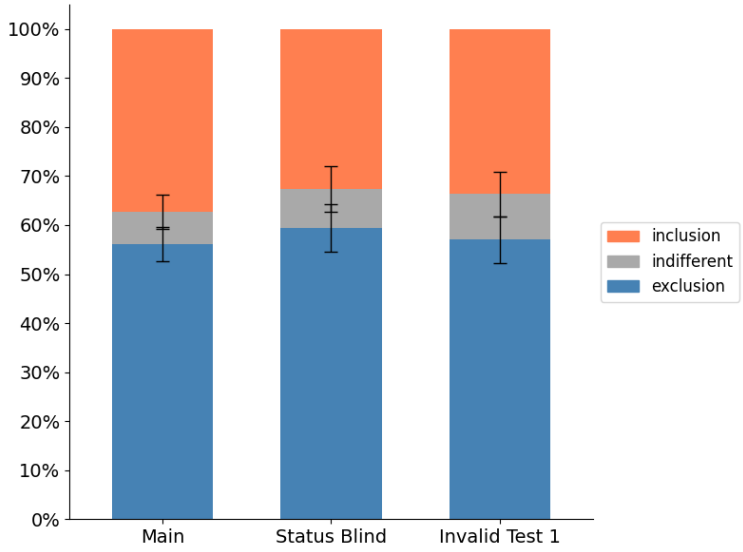

Figure 8: Information preferences across experiments: Invalid and Biased Prep scenario

Notes: This figure compares the proportions of spectators who prefer to include, exclude, or are indifferent about the inclusion of Test 2 scores between three treatments under the Invalid and Biased Prep scenario. Error bars represent the standard errors of the proportions of inclusion and exclusion preferences.

available student information, people tend to evaluate the scores in isolation when they consider their demand.

7 Attitudes toward Test-Blind and Test-Optional Admission Policies in the Real World

While we have identified procedural fairness and usefulness concerns as the two main drivers of test score exclusion in our controlled experiment, one remaining question is whether these considerations also influence people's attitudes toward test policies for real-world college admissions. To address this question, we survey spectators' attitudes toward test-blind and test-optional college admission policies at the end of the experiment and request open-ended rationales for their responses. By using GPT-3.5 to summarize these rationales, we find that fairness and usefulness concerns remain the primary considerations for supporting or opposing test-blind and test-optional policies (see Appendix C for details of the summarizing procedure). We then manually classify

each spectator's rationale in terms of whether it mentions these two concerns. 20% and 55% of rationales mention the fairness and usefulness of standardized test scores, respectively. Moreover, spectators who mention fairness in justifications of their information preferences in the experiment are also more likely to mention fairness in their rationales for attitudes toward real-world test policies ($r = 0.10$, $p = 0.01$). These results indicate that the considerations that drive information preferences in our experiment also influence attitudes toward test policies for college admissions.

Nevertheless, while supports for test-blind and test-optional admission policies are highly correlated with each other ($r = 0.70$, $p < 0.001$), they do not correlate with preferences for excluding Test 2 scores in the experiment (see Table A3). This is not surprising, given the random assignment of test prep scenarios in our experiment. For instance, a spectator randomly assigned to the Invalid (but unbiased) Prep scenario may not necessarily view the SAT and ACT as invalid and unbiased. Therefore, we should not expect their attitudes toward these standardized tests to align with their preferences regarding Test 2 scores. Indeed, spectators' self-reported rationales reveal various factors that shape their perceptions of standardized test scores' validity and bias. For example, in addition to test prep access (which is mentioned by 9% of spectators), spectators also mention differential test-taking skills (13%) and the across-applicant comparability of scores (7%) as factors that affect their attitudes toward test policies. Overall, while our study provides proof of concept that perceptions of a test's validity and bias affect whether people prefer to include it as an admission criterion, further research is needed to investigate the specific factors that shape these perceptions.

8 Discussion and Conclusion

In this paper, we examine how individuals' preferences to include or exclude test scores as an admission criterion are influenced by the validity and bias of these scores. We conduct an experiment where spectators decide which students from an introductory data analysis course are admitted to an advanced course. Before making admission decisions, spectators choose whether

to access students' scores from a test of the introductory course, which could have been influenced by test preparation. Our main finding indicates that the preference to exclude test scores becomes more prevalent when test preparation renders the scores invalid and/or biased. This preference is robust, remaining unaffected by the availability and quality of other student information, such as family income and other test scores. It also persists when admission decisions are made by others.

Using both direct evidence from self-reported reasoning and behavioral evidence, we show that the preference to exclude invalid and/or biased test scores is driven by procedural fairness concerns and a perception that the scores are not very useful for admission decisions. However, as spectators gain experience making admission decisions, the demand for test score information increases, especially among those who relied heavily on the scores during their decision-making process. This result demonstrates that people's demand for test score information can change as they learn about the usefulness of these scores through experience.

In this study, we focus on the attitudes of the general public. Laypeople's attitudes toward college admission policies are important because they shape the social pressure that influences college policymaking.[22] Nevertheless, it is also important to understand other factors that affect which test policy a college adopts. Colleges often aim to achieve complex objectives for their enrolled class (Butcher et al., 2013; Jaquette and Curs, 2015; Arcidiacono and Lovenheim, 2016) while facing pressures from donors (Golden, 2007) and even scrutiny from the authorities (Alfonseca, 2023). While these factors are certainly outside the scope of this paper, they represent fruitful avenues for future research.

By varying the availability and nature of test prep, our experiment identifies how the validity and bias of test scores affect people's attitudes toward their role in admissions. In Section 7, we show that these two properties are the main determinants of people's attitudes toward test policies in real-world college admissions. In Appendix D, we analyze newspaper articles to demonstrate that these two properties are also the primary considerations mentioned in media discussions on

[22]This influence can be direct, as some key laws governing admission policies, such as the Affirmative Action bans in California and Colorado, are determined through referendums. People can also exert influence indirectly through donations, advocacy, or protests.

test policies. Despite the centrality of the validity and bias of test scores in the public debate, there are other factors left out of our experiment that could affect public opinions on these policies. For example, not requiring standardized test scores may reduce the wasteful effort and costs students incur on test prep. In addition, eliminating these requirements may broaden the applicant pool (Saboe and Terrizzi, 2019). While these considerations are less frequently discussed than the focus of our study, they represent important subjects for future research.

Lastly, although this paper focuses on information preferences for admission decisions, the findings have potentially generalizable implications for other data-driven allocation problems. In hiring, loan approval, pre-trial release, and medical treatment allocations, decision-makers have potential access to various types of information, some of which may be inaccurate for or biased against certain groups. Our results imply that, due to procedural fairness concerns, decision-makers may choose to exclude these types of information even if they are useful for decision-making. This could explain the popular and, sometimes, legally required practice of eliminating inputs correlated with protected classes (such as race) when training predictive algorithms (Yang and Dobbie, 2020). While there is a large body of research on the legality and pragmatics of these practices, future research can investigate public attitudes on these issues.

References

AERA/APA/NCME. Standards for educational and psychological testing, 2014.

K. Alfonseca. Legacy college admissions under scrutiny after SCOTUS ruling. *ABC News*, jul 2023. URL `https://abcnews.go.com/US/legacy-college-admissions-scrutiny-scotus-ruling/story?id=100608150`. Accessed on August 5, 2023.

S. Ambuehl and S. Li. Belief updating and the demand for information. *Games and Economic Behavior*, 109:21–39, 2018.

P. Andre. Shallow meritocracy. *Review of Economic Studies*, page rdae040, 2024.

P. Arcidiacono and M. Lovenheim. Affirmative action and the quality–fit trade-off. *Journal of Economic Literature*, 54(1):3–51, 2016.

B. Bartling, E. Fehr, and H. Herz. The intrinsic value of decision rights. *Econometrica*, 82(6): 2005–2039, 2014.

V. Bartoš, M. Bauer, J. Chytilová, and F. Matějka. Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–1475, 2016.

P. Bhattacharya and J. Mollerstrom. Lucky to work. 2022.

J. A. Bohren, P. Hull, and A. Imas. Systemic discrimination: Theory and measurement. 2023.

G. E. Bolton, J. Brandts, and A. Ockenfels. Fair procedures: Evidence from games involving lotteries. *The Economic Journal*, 115(506):1054–1076, 2005.

E. Borghesan. The heterogeneous effects of changing sat requirements in admissions: An equilibrium evaluation. 2023.

K. F. Butcher, C. Kearns, and P. J. McEwan. Giving till it helps? alumnae giving and children's college options. *Research in Higher Education*, 54:499–513, 2013.

A. W. Cappelen, J. Mollerstrom, B.-A. Reme, and B. Tungodden. A meritocratic origin of egalitarian behaviour. *The Economic Journal*, 132(646):2101–2117, 2022.

A. W. Cappelen, C. Cappelen, and B. Tungodden. Second-best fairness: The trade-off between false positives and false negatives. *American Economic Review*, 113(9):2458–2485, 2023.

A. W. Cappelen, T. De Haan, and B. Tungodden. Fairness and limited information: Are people bayesian meritocrats? *Journal of Public Economics*, 233:105097, 2024.

A. Chakraborty and L. Henkel. The role of interpersonal uncertainty in prosocial behaviors. 2024.

J. Chan and E. Eyster. Does banning affirmative action lower college student quality? *American Economic Review*, 93(3):858–872, 2003.

G. Charness, R. Oprea, and S. Yuksel. How do people choose between biased information sources? evidence from a laboratory experiment. *Journal of the European Economic Association*, 19(3): 1656–1691, 2021.

R. Chetty, D. J. Deming, and J. N. Friedman. Diversifying society's leaders? the causal effects of admission to highly selective private colleges. Technical report, National Bureau of Economic Research, 2023.

K. Coffman, S. Kostyshak, and P. Saygin. Choosing and using information in evaluation decisions. 2024.

M. Conlin, S. Dickert-Conlin, and G. Chapman. Voluntary disclosure and the strategic behavior of colleges. *Journal of Economic Behavior & Organization*, 96:48–64, 2013.

W. Dessein, A. Frankel, and N. Kartik. Test-optional admissions. *arXiv preprint arXiv:2304.07551*, 2023.

S. Dynarski, A. Nurshatayeva, L. C. Page, and J. Scott-Clayton. Addressing nonfinancial barriers to college access and success: Evidence and policy implications. In *Handbook of the Economics of Education*, volume 6, pages 319–403. Elsevier, 2023.

C. L. Exley and K. Nielsen. The gender gap in confidence: Expected but not accounted for. *American Economic Review*, 114(3):851–885, 2024.

C. L. Exley, R. Fisman, J. B. Kessler, L.-P. Lepage, X. Li, C. Low, X. Shan, M. Toma, and B. Zafar. Information-optional policies and the gender concealment gap. Technical report, National Bureau of Economic Research, 2024.

P. Eyal, R. David, G. Andrew, E. Zak, and D. Ekaterina. Data quality of platforms and panels for online behavioral research. *Behavior research methods*, pages 1–20, 2021.

A. Falk, S. Heuser, and D. Huffman. Moral luck: Mechanisms, robustness, and prevalence. 2023.

S. Fath, R. P. Larrick, and J. B. Soll. Blinding curiosity: Exploring preferences for "blinding" one's own judgment. *Organizational Behavior and Human Decision Processes*, 170:104135, 2022.

S. Fath, R. P. Larrick, and J. B. Soll. Encouraging self-blinding in hiring. *Behavioral Science & Policy*, 9(1):45–57, 2023.

H. Feder and A. Bello. Why college admissions should remain test optional/test free (despite what the new york times says), March 2024. URL `https://fairtest.org/wp-content/uploads/2024/03/TestOptionalReportFinal.pdf`. Accessed: 2024-07-04.

D. Fudenberg, P. Strack, and T. Strzalecki. Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–3684, 2018.

N. Garg, H. Li, and F. Monachou. Dropping standardized testing for admissions trades off information and access, 2023. URL `https://arxiv.org/abs/2010.04396`.

D. Golden. *The Price of Admission (Updated Edition): How America's Ruling Class Buys Its Way into Elite Colleges–and Who Gets Left Outside the Gates*. Crown, 2007.

R. Golman, D. Hagmann, and G. Loewenstein. Information avoidance. *Journal of economic literature*, 55(1):96–135, 2017.

M. Guan. Choosing between information bundles. 2023.

M. Guan, R. Oprea, and S. Yuksel. Too much information. 2023.

M. Y. Gurdal, J. B. Miller, and A. Rustichini. Why blame? *Journal of Political Economy*, 121(6): 1205–1247, 2013.

Y. Halevy, D. Walker-Jones, and L. Zrill. *Difficult decisions*. University of Toronto, Department of Economics, 2023.

O. Jaquette and B. R. Curs. Creating the out-of-state university: Do public universities increase nonresident freshman enrollment in response to declining state appropriations? *Research in Higher Education*, 56:535–565, 2015.

Y. Liang. Boundedly rational information demand. Technical report, 2023.

E. S. Phelps. The statistical theory of racism and sexism. *The American Economic Review*, 62(4): 659–661, 1972.

A. Rambachan, J. Kleinberg, S. Mullainathan, and J. Ludwig. An economic approach to regulating algorithms. Technical report, National Bureau of Economic Research, 2020.

M. Saboe and S. Terrizzi. Sat optional policies: Do they influence graduate quality, selectivity or diversity? *Economics Letters*, 174:13–17, 2019.

P. R. Sackett, N. R. Kuncel, A. S. Beatty, J. L. Rigdon, W. Shen, and T. B. Kiger. The role of socioeconomic status in sat-grade relationships and in college admissions decisions. *Psychological science*, 23(9):1000–1007, 2012.

S. T. Trautmann. Procedural fairness and equality of opportunity. *Journal of Economic Surveys*, 37(5):1697–1714, 2023.

S. E. Woo, J. M. LeBreton, M. G. Keith, and L. Tay. Bias, fairness, and validity in graduate-school admissions: A psychometric perspective. *Perspectives on Psychological Science*, 18(1):3–31, 2023.

C. S. Yang and W. Dobbie. Equal protection under algorithms: A new statistical and legal framework. *Michigan Law Review*, pages 291–395, 2020.

Appendix

A Additional Figures and Tables

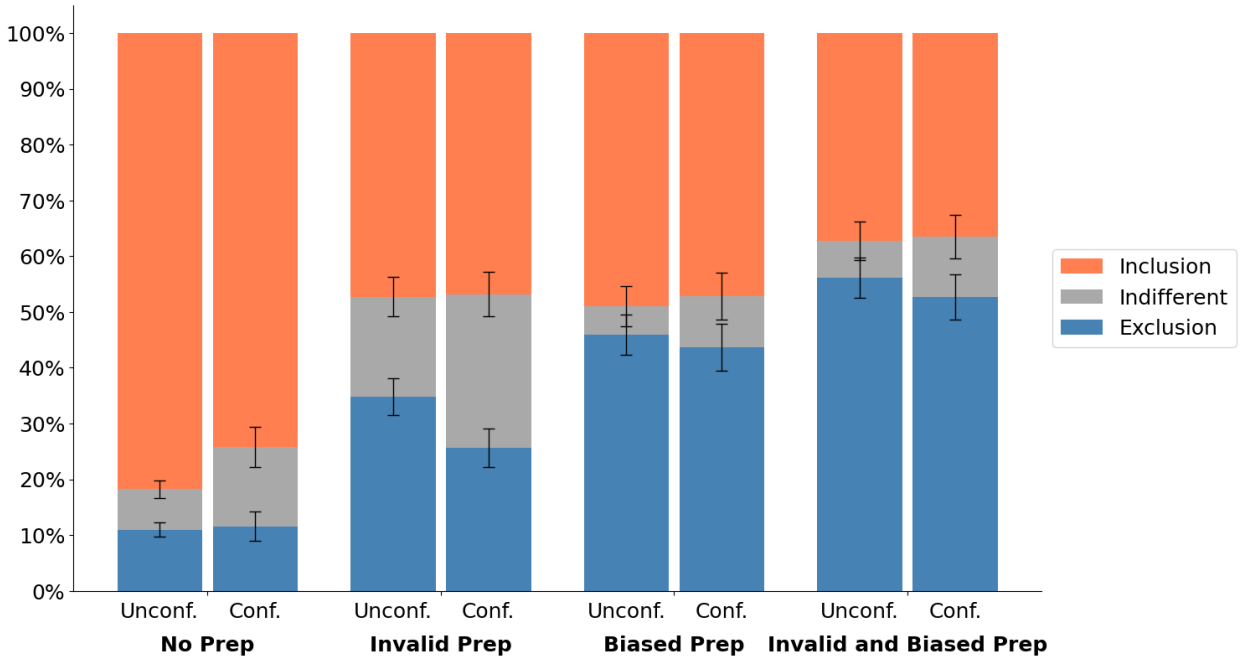

Figure A1: Confirmed Information preferences across test prep scenarios (before admission experience)

Notes: This figure compares the distributions of information preferences for the true scenario before and after confirmation. The unconfirmed preferences, represented by the left bars, are classified solely based on the spectators' reported preferences before admission experience. After the true scenario is realized, spectators who expressed a preference for including or excluding Test 2 scores under this scenario are asked to complete a real-effort task to confirm their preferences. If they choose not to confirm, their confirmed preferences (represented by the right bars) will be classified as indifferent. Error bars represent the standard errors of the proportions of inclusion and exclusion preferences.

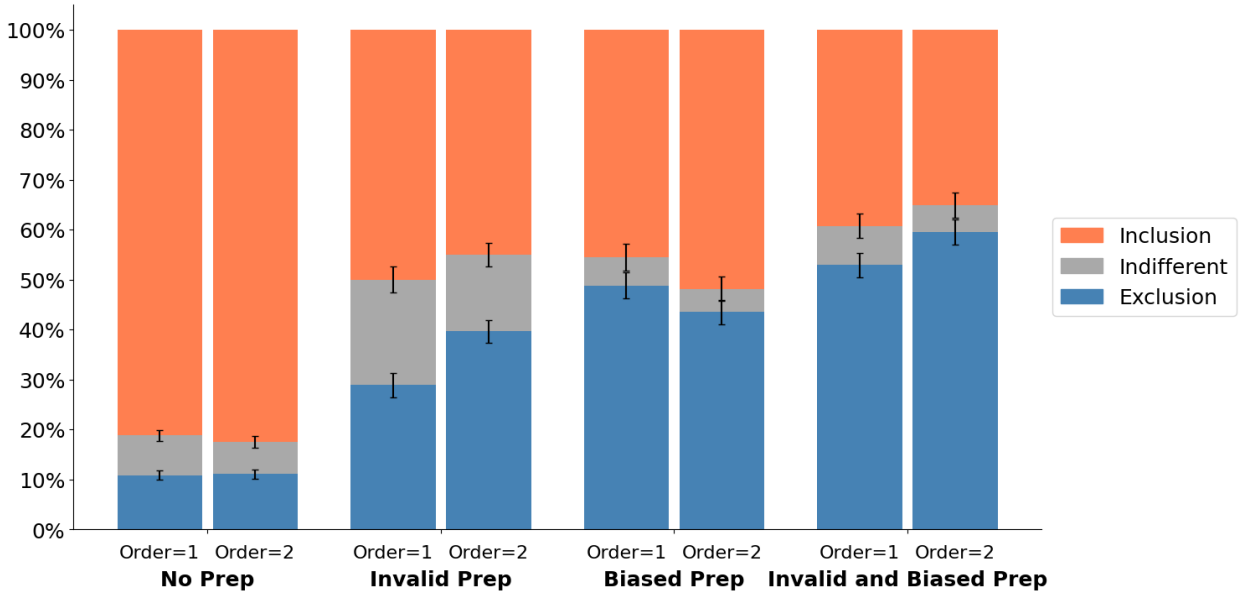

Figure A2: Information preferences by the order between test prep scenarios (before admission experience)

Notes: This figure shows results on the order effect in information preferences. For each scenario, the left (right) bar shows the distribution of information preferences when the scenario is presented first (second). Error bars represent the standard errors of the proportions of inclusion and exclusion preferences.

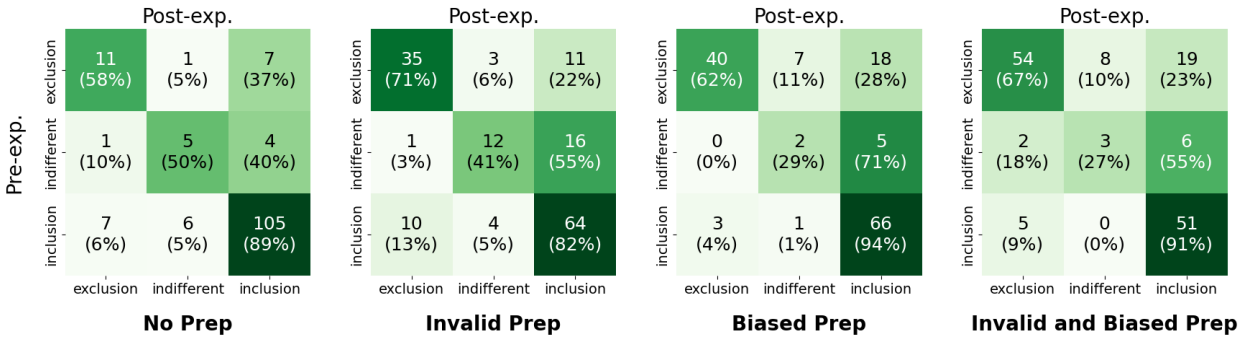

Figure A3: Transition matrix for information preferences before and after admission experience

Notes: This figure shows the numbers of spectators in cells defined by their information preferences before admission experience and after admission experience. Numbers in brackets represent proportions relative to the number of spectators with the same information preference before admission experience.

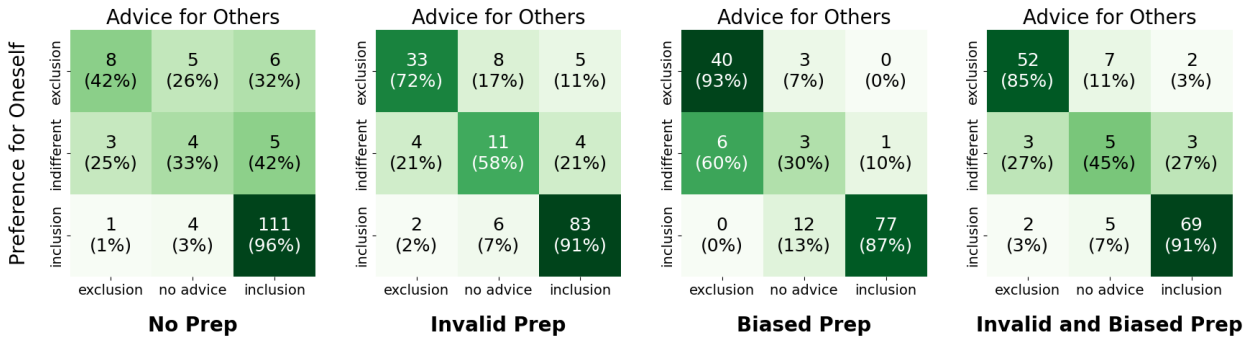

Figure A4: Transition matrix for information preferences for oneself and advice for others across test prep scenarios

Notes: This figure shows the joint distributions of information preferences for oneself (after admission experience) and advice to others for each test prep scenario. Numbers in brackets represent proportions relative to the number of spectators with the same information preference.

	Information Preferences	
	(1)	(2)
Test prep scenario (ref: No Prep)		
Invalid Prep	-0.520***	-0.601***
	(0.075)	(0.071)
Biased Prep	-0.687***	-0.633***
	(0.089)	(0.081)
Invalid & Biased Prep	-0.950***	-0.908***
	(0.081)	(0.076)
Test prep scenario order	-0.022	-0.027
	(0.048)	(0.050)
Age bracket		0.054**
		(0.023)
Gender = Male		0.008
		(0.054)
Race = White		0.058
		(0.060)
Employment status (ref: Working full-time)		
Working part-time		0.070
		(0.072)
Unemployed and looking for work		0.104
		(0.099)
Homemaker or stay-at-home parent		0.038
		(0.101)
Student		0.100
		(0.143)
Retired		-0.038
		(0.117)
Education Level		0.030
		(0.020)
Liberal orientation		0.016
		(0.022)
Income bracket		0.023
		(0.018)
Spectator FE	Yes	No
R^2	0.58	0.18
Observations	1,186	1,084

Table A1: Regression analysis of the effects of test prep scenarios on information preferences

Notes: This table shows OLS estimates of the effects of test prep scenarios on information preferences. Information preferences are coded as follows: include: 1; indifferent: 0; exclude: -1. Test prep scenario order is 1 if the scenario is presented first and 2 otherwise. Age bracket is coded as follows: Under 18: 1; 18-24 years old: 2; 25-34 years old: 3; 35-44 years old: 4; 45-54 years old: 5; 55-64 years old: 6; 65+ years old: 7. Education level is coded as follows: Some high school or less: 1; High school diploma or GED: 2; Some college, but no degree: 3; Associates or technical degree: 4; Bachelor's degree: 5; Graduate or professional degree: 6. Income bracket is coded as follows: Less than $25,000: 1; $25,000-$49,999: 2; $50,000-$74,999: 3; $75,000-$99,999: 4; $100,000-$149,999: 5; $150,000 or more: 6. Liberal leaning is coded as follows: political leaning is very conservative: -2; somewhat conservative: -1; neither liberal nor conservative: 0; somewhat liberal: 1; very liberal: 2. Numbers in parentheses are robust standard errors clustered at the participant level. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

	Average decision time when Test 2 scores are revealed				Average additional decision time when Test 2 scores are revealed			
	No Prep	Invalid Prep	Biased Prep	Invalid and Biased Prep	No Prep	Invalid Prep	Biased Prep	Invalid and Biased Prep
Pre-experience info preference	-3.06	-5.60	4.66	-0.99	0.80	-5.13	-14.34*	2.72
	(1.88)	(3.75)	(2.98)	(3.29)	(1.63)	(3.91)	(6.47)	(2.81)
Post-experience info preference	1.25	3.67	-2.28	3.50	2.30	-1.38	18.04*	-0.75
	(1.88)	(3.71)	(3.20)	(3.27)	(1.62)	(3.87)	(6.94)	(2.79)
R^2	0.02	0.01	0.02	0.01	0.03	0.02	0.05	0.01
Observations	147	156	142	147	147	156	142	147

Table A2: Decision time and information preferences by scenario

Notes: This table shows the OLS estimates of regressing decision time on information preferences under each test prep scenario. The dependent variable of columns 1 to 4 is the average decision time for Groups 4 to 6 where Test 2 scores are revealed. The dependent variable of columns 5 to 8 is the average decision time for Groups 4 to 6 minus the average decision time for Groups 1 to 3 where Test 2 scores are not revealed. The independent variables, information preferences before and after admission experience, are both coded as follows: include: 1; indifferent: 0; exclude: -1. Numbers in brackets are standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A3: Information preferences and attitudes towards test policies

	Information Preferences (after admission experience)			
	No Prep	Invalid Prep	Biased Prep	Invalid and Biased Prep
Support test-blind policies	-0.09	-0.01	-0.16	0.08
	(0.28)	(0.93)	(0.06)	(0.36)
Support test-optional policies	-0.11	-0.03	-0.04	0.10
	(0.19)	(0.69)	(0.63)	(0.21)

Notes: This table shows the correlations between support for test-blind or test-optional college admission policies (relative to test-required policies) and information preferences (after admission experience) for each test prep scenario. Information preferences are coded as follows: include: 1; indifferent: 0; exclude: -1. Support for test policies are coded as follows: strongly support test-required policies: -2; somewhat support test-required policies: -1; indifferent: 0; somewhat support test-blind/optional policies: 1; strongly support test-blind/optional policies: 2. Numbers in brackets are p-values for correlation coefficients.

B Procedure for Summarizing Rationales for Test 2 Scores Exclusion and Inclusion

We use large language models to summarize the main rationales for Test 2 score information preferences. This approach allows us to avoid predefining rationale categories, which could potentially bias our analysis. The summarizing procedure is as follows. First, we randomly partition all 593 open-ended justifications for preferences over including or excluding Test 2 scores from across all test prep scenarios into sets of 30 justifications (one set has 23 justifications). This leads to 20 sets of justifications. Second, for each set of justifications, we send Prompt 1 (copied below) to GPT-3.5-turbo to identify the top three mentioned rationales. This results in 20 lists of top three rationales. Third, we combines these lists with Prompt 2 (copied below) to summarize the overall top three rationales. We repeat this three-step process 100 times, each time generating a different partition of all the justifications. This leads to 100 outputs of GPT-identified top three rationales. Finally, we use Prompt 2 again to summarize these 100 outputs.

The resulting summary of the three most common reasons for excluding Test 2 scores are:

1. Concerns about fairness and bias, especially if only certain students had access to test preparation.

2. Belief that test preparation may provide an unfair advantage and inflate scores, compromising the credibility of the evaluation.

3. Doubts about the accuracy and relevance of Test 2 scores in reflecting students' true abilities.

The top three rationales for including Test 2 scores:

1. To assess the impact of test preparation on students' performance and to understand improvements or declines in scores.

2. Desire for more information to make a comprehensive assessment of students' performance and to compare outcomes with and without test preparation.

3. To have a more comprehensive understanding of students' grasp of the material and performance.

From this summary, we conclude that the main reason for excluding Test 2 scores are concerns about the fairness and usefulness of the scores whereas the main reason for including Test 2 scores is that they are useful for understanding students' skills.

Prompt 1: Here are some open-ended responses from a survey asking people whether they would like to see a test's score and use it to evaluate students. Each open-ended response is separated by the semicolon. If the open-ended response is 'this person didn't write anything', you can ignore it because the participant didn't provide any reasoning. The context is as follows: students take two tests about data analysis, Test 1 and Test 2. Test 1 score for each student will always be available and people decide whether they would like to see students' Test 2 score. Some students receive test preparation for Test 2 and the test preparation could boost students' scores. For the test preparation, there could be three scenarios and each survey participants read one scenario: First scenario, the test preparation is only available to high-income students, and the test preparation provides one question's answer in Test 2 and there are 5 questions in Test 2. This test preparation cannot improve students' skills in data analysis. Second scenario, the test preparation is available to both high and low-income students, and the test preparation provides one question's answer in Test 2 and there are 5 questions in Test 2. This test preparation cannot improve students' skills in data analysis. Third scenario, the test preparation is only available to high-income students, and the test preparation provides general training to improve students' performance in the test, and such general training could improve students' data analysis skills. Survey participants provide their open-ended responses to explain why they want or they don't want to see Test 2 scores when evaluating students. Although each participant knows which group(s) receive the test preparation and what the test preparation is, to make sure your coding only depends on participants' open-ended response, you will not see which scenario is, although some participants may elaborate the scenarios in their open-ended responses. Your task is to summarize the most common 3 reasons why to include or exclude Test 2 scores from these open-ended responses. Please notice your sum-

mary should focus on the reason people provided about why they want to include or exclude Test 2 scores, instead of the decision to include or exclude scores. You should not include any personal opinions or interpretations in your summary, but rather focus on objectively presenting the reasons from the open-ended responses.

Prompt 2: Here are some summaries and each summary is separate by the semicolon. The summary might begin with something like 'the most common reasonings from the open-ended responses are' and the most common 3 topics are listed afterwards. Each summary extracts the most common topics from some participants' open-ended responses, for example, the first summary is the summary of most common topics from the first 30 participants' responses, and the second summary is for the 31st to the 60th participant's response, so on and so forth. Your task is to review all summaries and extract the most common 3 reasons why to include or exclude Test 2 scores from these open-ended responses. Please notice your extraction should focus on the reason people provided about why they want to include or exclude Test 2 scores, instead of the decision to include or exclude scores. You should not include any personal opinions or interpretations in your summary, but rather focus on objectively presenting the reasonings from the open-ended responses.

C Procedure for Summarizing Rationales for Attitudes towards Test-Blind/Optional policies

We use the following procedure to summarize the main rationales for attitudes towards test-blind and test-optional policies for real-world college admissions. First, we randomly divide all 593 spectators in the main treatment into groups of 30 (one group has 23 spectators). This leads to 20 groups. Second, for each group, we combine the spectators' justifications for their attitudes toward these two test policies and Prompt 3 (copied below), and then send it to GPT-3.5-turbo to identify the top three mentioned rationales. This results in 20 lists of top three rationales. Third, we combines these lists with Prompt 4 (copied below) to summarize the overall top three rationales. We repeat this three-step process 100 times, each time generating a different partition of spectator

groups. This leads to 100 outputs of GPT-identified top three rationales. Due to token length constraints, we split the 100 outputs into two halves. We then use Prompt 4 to summarize each half separately. Finally, we apply Prompt 4 again to distill the three most common rationales from the combined summaries of both halves.

The resulting summary of the three most common reasons for supporting the test-blind and test-optional policies from the open-ended responses are:

1. Concerns about fairness, equity, and inclusivity in the admissions process
2. Belief that standardized tests may not accurately reflect a student's abilities or potential for success in college
3. Preference for a holistic evaluation of applicants beyond just test scores.

The top three rationales for for opposing the test-blind and test-optional policies are:

1. Belief in the importance of standardized tests as reliable indicators of academic readiness
2. skepticism about alternative evaluation methods
3. Concerns about the potential consequences of not considering test scores.

From this summary, we conclude that the main reason for supporting test-blind/optional policies are concerns about the fairness and usefulness of the scores, as well as support of holistic evaluation whereas the main reason for opposing test-blind/optional policies is that standardized test scores are relatively useful measures for students' skills and knowledge.

Prompt 3: Here are some open-ended responses from a survey asking people whether they support or not support the test-blind policy and the test-optional policy. Each open-ended response is separated by the semicolon. If the open-ended response is 'this person didn't write anything', you can ignore it because the participant didn't provide any reasoning. Your task is to summarize the top 3 common reasons for supporting or opposing test-blind and test-optional policies from the provided open-ended responses. In your summary, you may combine the reasons for support

or oppose the test-blind policy and the test-optional policy together. You should not include any personal opinions or interpretations in your summary, but rather focus on objectively presenting the reasoning from the open-ended responses.

Prompt 4: Here are some summaries and each summary is separate by the semicolon. The summary might begin with something like 'the most common reasoning from the open-ended responses are' and the most common 3 topics are listed afterwards. Each summary extracts the most common topics from some participants' open-ended responses, for example, the first summary is the summary of most common topics from the first 30 participants' responses, and the second summary is for the 31st to the 60th participant's response, so on and so forth. Your task is to review all summaries and extract the most common 3 reasons why support or oppose the test-blind policy and the test-optional policy from these open-ended responses. You should not include any personal opinions or interpretations in your summary, but rather focus on objectively presenting the reasonings from the open-ended responses.

D Summarizing Rationales for Opinions on College Admission Test Policies in Media Article

To further validate the central role that validity and bias play in public attitudes toward college admission test policies, we construct a dataset of media articles discussing these policies and analyze the main arguments therein using a large language model.

Specifically, we first use "test optional", "test blind", "reinstate standardized test" as keywords to search within the top 10 most circulated US newspapers [23], including The New York Times, The Wall Street Journal, The Washington Post, USA Today, and The Chicago Tribune, along with The Boston Globe, The Star Tribune, The New York Post, Newsday, dating back to 2000. This search results in a total of 172 news articles.

[23]Majid, Aisha (April 6, 2023). "Mail joins 100k Club: Exclusive ranking of world's top paywalled news publishers". Press Gazette. https://pressgazette.co.uk/paywalls/digital-news-subscriptions-ranking-2023/, Retrieved November 6, 2023.

Then, we use the following procedure tosummarize these 172 news articles. First, we send Prompt 5 (copied below) to GPT-3.5-turbo to summarize each article's main points. Next, we randomly partition all 172 news article summary into sets of 30 summaries (one set has 22 summaries). This leads to 6 sets of summaries. Second, for each set of summaries, we send Prompt 6 (copied below) to GPT-3.5-turbo to identify the top three mentioned rationales for supporting or opposing test-blind/test-optional policies. This results in 20 lists of top three rationales. Third, we combine these lists with Prompt 7 (copied below) to summarize the overall top three rationales. We repeat this three-step process 100 times, each time generating a different partition of all the justifications. This leads to 100 outputs of GPT-identified top three rationales. Finally, we use Prompt 7 again to summarize these 100 outputs.

The top three rationales for test-blind/test-optional polices, or reinstatement of test requirement are:

1. Concerns about bias, lack of predictive value, and inequities in standardized testing, leading to a push for more equitable assessments and promoting diversity and inclusivity in admissions.

2. Emphasis on holistic evaluation of applicants beyond standardized test scores, considering various factors like high school grades, coursework rigor, and extracurricular activities to provide a more comprehensive view of student potential.

3. Addressing equity and access in college admissions, including concerns about disparities faced by underrepresented and disadvantaged students, leading to a reevaluation of standardized test requirements to ensure fairness and inclusivity in the admissions process.

From this summary, we conclude that the validity and bias of standardized test scores are the central concerns in media discussion about college admission test policies.

Prompt 5: I want you to act as a news article summarizer. I will provide you with a news article about test-optional or test-blind college admissions policies, and you will create a summary of the main points and findings of the news article. Your summary should be concise and should

accurately and objectively communicate the key points of the news article. You should not include any personal opinions or interpretations in your summary, but rather focus on objectively presenting the information from the news article. Your summary should be written in your own words and should not include any direct quotes from the news article. Please ensure that your summary is clear, concise, and accurately reflects the content of the original news article. Please only use one sentence to summarize.

Prompt 6: Below is a list of summarized news articles, and each summary is separated by a semicolon. Each summary is a summary of one news article reporting the rationales for test-optional or test-blind college admissions policies, or reinstating test requirements. Your task is to find the top 3 commonly discussed rationales from these news articles.

Prompt 7: Below is a list of summarized topics from news articles, and each summarized news article topic is separated by a semicolon. Each summary is a summary of ten news article reporting the rationales for test-optional or test-blind college admissions policies, or reinstating test requirements. Each summary of commonly discussed rationales might start with 'the top 3 commonly discussed rationales from these news articles are' and then three rationales are listed. Your task is to find the top 3 commonly discussed rationales from these summaries.

E Survey for the Study

Introduction to study

Welcome to our survey! Your answer to the survey is important – it helps our study and may directly impact others – so please answer carefully.

Please read the survey instructions carefully. All content of the instructions is true, as guaranteed by the CMU Institutional Review Board (IRB).

We will check your understanding of the instructions at several points during the survey. You can proceed in the survey only after you pass our understanding check.

Instruction for the main treatment

Context

We taught some college students a basic course on data analysis. Once the course was finished, the students took two tests, each focusing on different things they learned.

In this survey, we will show you seven groups of students, with each group having 8 students. Your task is to admit 3 students from each group to a more advanced data analysis course.

Only one of these groups is made up of actual students who took our basic course, while the rest are fictional. Your decision to admit students from the real group will have some chance of being put into action. However, you will not be told which group is the real one, so all of the admission decisions you make are important.

Report Card

To help you decide which students to admit, we will give you a report card for each student. The report card will always include the student's **family income** level and **Test 1 score**. It may also include the student's **Test 2 score**.

About family income:
- Students with family incomes over $100,000 are classified as Higher-Income
- Students with family incomes under $100,000 are classified as Lower-Income

About Test 1:
- Number of questions: 5
- Maximum score: 10 points
- Higher-Income and Lower-Income students performed similarly in Test 1.
- No student received any test prep for Test 1.

About Test 2:
- Number of questions: 5
- Maximum score: 10 points
- When there is no test prep, Higher-Income and Lower-Income students usually perform similarly in Test 2.
- Some students may have received test prep for Test 2.

We will tell you more about the test prep situation for Test 2, and then ask if you want us to include the Test 2 scores in the report cards.

But before that, you need to answer some questions to make sure that you understand the context. When you are

ready, please click Next. Once you click Next, you can no longer return to this page.

Instruction for the family income blind treatment

Context

We taught some college students a basic course on data analysis. Once the course was finished, the students took two tests, each focusing on different things they learned.

In this survey, we will show you seven groups of students. Each group has 4 students from higher-income families (income over $100,000) and 4 from lower-income families (income below $100,000). Your task is to admit 3 students from each group of 8 to a more advanced data analysis course.

Only one of these groups is made up of actual students who took our basic course, while the rest are fictional. Your decision to admit students from the real group will have some chance of being put into action. However, you will not be told which group is the real one, so all of the admission decisions you make are important.

Report Card

To help you decide which students to admit, we will give you a report card for each student. The report card will always include the student's Test 1 score but not the family income level. It may also include the student's Test 2 score.

About Test 1:

- Number of questions: 5
- Maximum score: 10 points
- Higher-Income and Lower-Income students performed similarly in Test 1.
- No student received any test prep for Test 1.

About Test 2:

- Number of questions: 5
- Maximum score: 10 points
- When there is no test prep, Higher-Income and Lower-Income students usually perform similarly in Test 2.
- Some students may have received test prep for Test 2.

We will tell you more about the test prep situation for Test 2, and then ask if you want us to include the Test 2 scores in the report cards.

But before that, you need to answer some questions to make sure that you understand the context. When you are ready, please click Next. Once you click Next, you can no longer return to this page.

Instruction for prediction treatment

Context

Some college students took two courses on data analysis, one basic and the other more advanced. They took two tests after finishing the basic course, Test 1 and Test 2, each focusing on different things they learned. For the advanced course, we also tested their performance.

In this survey, we will show you seven groups of students, with each group having 8 students. Your task is to guess which three students in each group performed the best in the advanced course.

Only one of these groups is made up of actual students who took our courses, while the rest are fictional. Your guess for the real group will determine your bonus – for each

correct guess, you will receive an additional $1 bonus. However, you will not be told which group is the real one, so all guesses are important.

The students won't know or be affected by your guesses in any way.

Report Card

To help you guess the top performers in the advanced course, we will give you a report card for each student. The report card will always include the student's **family income level** and **Test 1 score for the basic course**. It may also include the student's **Test 2 score for the basic course**.

About performance in the advanced course:
- Every student is evaluated on equal footing.

About family income:
- Students with family incomes over $100,000 are classified as Higher-Income
- Students with family incomes under $100,000 are classified as Lower-Income

About Test 1:
- Number of questions: 5
- Maximum score: 10 points
- Higher-Income and Lower-Income students performed similarly in Test 1.
- No student received any test prep for Test 1.

About Test 2:
- Number of questions: 5
- Maximum score: 10 points
- When there is no test prep, Higher-Income and Lower-Income students usually perform similarly in Test 2.
- Some students may have received test prep for Test 2.

We will tell you more about the test prep situation for Test 2, and then ask if you want us to include the Test 2 scores in the report cards.

But before that, you need to answer some questions to make sure that you understand the context. When you are ready, please click Next. Once you click Next, you can no longer return to this page.

Instruction for imperfect Test 1 treatment

Context

We taught some college students a basic course on data analysis. Once the course was finished, the students took two tests, each focusing on different things they learned.

In this survey, we will show you seven groups of students, with each group having 8 students. Your task is to admit 3 students from each group to a more advanced data analysis course.

Only one of these groups is made up of actual students who took our basic course, while the rest are fictional. Your decision to admit students from the real group will have some chance of being put into action. However, you will not be told which group is the real one, so all of the admission decisions you make are important.

Report Card

To help you decide which students to admit, we will give you a report card for each student. The report card will

always include the student's **family income** level and **Test 1 score**. It may also include the student's **Test 2 score**.

About family income:
- Students with family incomes over $100,000 are classified as Higher-Income
- Students with family incomes under $100,000 are classified as Lower-Income

About Test 1:
- Number of questions: 5
- Maximum score: 10 points
- Higher-Income and Lower-Income students performed similarly in Test 1.
- Some students were graded more leniently than others, but you won't know who they are.

About Test 2:
- Number of questions: 5
- Maximum score: 10 points
- All students were graded by the same standard, but some students may have received test prep for Test 2.
- When there is no test prep, Higher-Income and Lower-Income students usually perform similarly in Test 2.

We will tell you more about the test prep situation for Test 2, and then ask if you want us to include the Test 2 scores in

the report cards.

But before that, you need to answer some questions to make sure that you understand the context. When you are ready, please click Next. Once you click Next, you can no longer return to this page.

Comprehension check

Will the report card include each student's family income level?

- ○ Yes
- ○ No

Will the report card include each student's score in Test 1?

- ○ Yes
- ○ No

Did any student receive test prep for Test 1?

- ○ Yes

- ○ No

Which group performed better in Test 1?

- ○ Higher-Income students
- ○ Lower-Income students
- ○ Both groups performed similarly

When there is no test prep, which group usually perform better in Test 2?

- ○ Higher-Income students
- ○ Lower-Income students
- ○ Both groups performed similarly

Transition to information preference

You have correctly answered all understanding questions.

Now we will show you two scenarios about test prep for Test 2. For each scenario, we will ask if you want Test 2 scores to appear on the students' report cards.

One of these scenarios is real, and your choice for that one will determine the content of the report cards. So, please think carefully before choosing.

Information preference: invalid and biased prep

Scenario

Recall that the students' report cards always include their family income levels and Test 1 scores.

About Test 2, suppose that

- Only Higher-Income students received test prep.
- The prep gave students the answer for a random question in Test 2 in advance. The other four questions are not affected by it. The prep may have boosted their Test 2 scores by up to 2 points, but it didn't make them any better at data analysis.

Do you want students' Test 2 scores to appear on their report cards, in addition to family income levels and Test 1 scores?

○ Yes

- ○ No
- ○ I am indifferent

Can you explain your reasoning for this answer?

Information preference: biased prep

Scenario

Recall that the students' report cards always include their family income levels and Test 1 scores.

About Test 2, suppose that

- Only Higher-Income students received test prep.
- The prep gave students an additional insight that could be used in a question in Test 2. The other four questions are not affected by it. The prep may have boosted their Test 2

scores by up to 2 points and made them better at data analysis.

Do you want students' Test 2 scores to appear on their report cards, in addition to family income levels and Test 1 scores?

- ○ Yes
- ○ No
- ○ I am indifferent

Can you explain your reasoning for this answer?

Information preference: invalid prep

Scenario

Recall that the students' report cards always include their

family income levels and Test 1 scores.

About Test 2, suppose that

- Every student received test prep.
- The prep gave students the answer for a random question in Test 2 in advance. The other four questions are not affected by it. The prep may have boosted their Test 2 scores by up to 2 points, but it didn't make them any better at data analysis.

Do you want students' Test 2 scores to appear on their report cards, in addition to family income levels and Test 1 scores?

- ○ Yes
- ○ No
- ○ I am indifferent

Can you explain your reasoning for this answer?

Information preference: no prep

Scenario

Recall that the students' report cards always include their family income levels and Test 1 scores.

About Test 2, suppose that:

No student received any test prep.

Do you want students' Test 2 scores to appear on their report cards, in addition to family income levels and Test 1 scores?

- ○ Yes
- ○ No
- ○ I am indifferent

Can you explain your reasoning for this answer?

Reveal true scenario for admissions

Thank you for your answers.

For the groups of students you will make admission decisions for, Scenario [1] correctly describes the test prep situation for Test 2. That is,

- No student received any test prep.

Please keep this in mind when you make admission decisions. We will test your understanding of this scenario on the next few pages.

Previously, you indicated that you want Test 2 scores to appear on students' report cards in this scenario. If you want to confirm this answer, please type the sentence from the black box into the text box below.

Once we receive your confirmation, we will make sure that for the majority of student groups, the report cards will display the scores for Test 2. Alternatively, if you leave the text box empty, there might be fewer situations where Test 2 scores are included.

I want Test 2 scores to be on the report cards.

Thank you for your answer.

Before making admission decisions, you need to answer

some questions to make sure you understand the test prep situation. Please click Next when you are ready.

Comprehension checks

Did any student receive test prep for Test 2?

- ○ No
- ○ Only Higher-Income students did
- ○ Only Lower-Income students did
- ○ All students did

What did the test prep for Test 2 entail?

- ○ It revealed the answer for a random question in Test 2 in advance.
- ○ It provided an additional insight that could be used in a question in Test 2.

Could the test prep help boost students' test scores?

- ○ Yes, by up to 2 points
- ○ No

Could the test prep make students better at data analysis?

- ○ Yes
- ○ No

Transition to admissions tasks without Test 2

Now, we will ask you to make admission decisions for three groups of students. The report cards for these students will only include their family income levels and Test 1 scores, but not their Test 2 scores.

Remember:
- No student received test prep for either test.
- When there is no test prep, Higher-Income and Lower-Income students perform similarly on both tests.

If you are ready, please click Next.

Example: admissions without Test 2

Round 1

In what order do you want us to show you the report cards of the 8 students?

Group by income level, then sort by Test 1 scores within each group

Sort all students by Test 1 scores

Below are the report cards of 8 students. Please admit 3 students to the advanced data analysis course. If you can't select exactly 3 students, please select the option "I can't decide which 3 students to admit, in which case some other study participant will decide whom to admit.

- [ ] Higher-Income group, Test 1 score: 9
- [ ] Higher-Income group, Test 1 score: 8
- [ ] Higher-Income group, Test 1 score: 5
- [ ] Higher-Income group, Test 1 score: 4
- [ ] Lower-Income group, Test 1 score: 9
- [ ] Lower-Income group, Test 1 score: 7
- [ ] Lower-Income group, Test 1 score: 6
- [ ] Lower-Income group, Test 1 score: 4
- [ ] I cannot decide which 3 students to admit.

Submit my selection

Transition to admissions tasks with Test 2

Thank you for your answers. In the next part, we will ask you to make admission decisions for three more groups of students. The main difference from the previous decisions is that in addition to the students' family income levels and Test 1 scores, the report cards will also include their Test 2 scores.

Remember:
- No student received test prep for either test.
- When there is no test prep, Higher-Income and Lower-Income perform similarly on both tests.

If you are ready, please click Next.

Example: admissions with Test 2

Round 6

In what order do you want us to show you the report cards of the 8 students?

Group by income level. Within each group, sort by Test 1 and then Test 2

Sort all students by Test 1 and then Test 2

Group by income level. Within each group, sort by Test 2 and then Test 1

Sort all students by Test 2 and then Test 1

Below are the report cards of 8 students. Please admit 3 students to the advanced data analysis course. If you can't select exactly 3 students, please select the option "I can't decide which 3 students to admit," in which case some other study participant will decide whom to admit.

- [ ] Higher-Income group, Test 1 score: 7, Test 2 score: 9
- [ ] Higher-Income group, Test 1 score: 7, Test 2 score: 8
- [ ] Higher-Income group, Test 1 score: 7, Test 2 score: 5
- [ ] Higher-Income group, Test 1 score: 6, Test 2 score: 5
- [ ] Lower-Income group, Test 1 score: 7, Test 2 score: 9
- [ ] Lower-Income group, Test 1 score: 7, Test 2 score: 7
- [ ] Lower-Income group, Test 1 score: 7, Test 2 score: 5
- [ ] Lower-Income group, Test 1 score: 7, Test 2 score: 4
- [ ] I cannot decide which 3 students to admit.

Submit my selection

Information preference: after admissions experience

Thanks for your answers. You have one more admission decision to make.

Before that, we need to ask: do you want Test 2 scores on the students' report cards? You've answered this before. You can stick with that answer or change it. Both your previous and current answers have a 50% chance of deciding the report card content for this final decision.

Do you want students' Test 2 scores to appear on their report cards, in addition to family income levels and Test 1 scores?

- ○ Yes
- ○ No
- ○ I'm indifferent

Information preference: advice for others

Thank you for your answer. Now, we'd appreciate your advice for future participants in this study on whether to include Test 2 scores on the report cards. They will face the same test prep situation as you. That is,

- No student received test prep for either test.
- When there is no test prep, Higher-Income and Lower-Income perform similarly on both tests.

Would you advise other participants to include Test 2 scores in students' report cards? Some participants may see your advice.

- ○ Yes, I advise them to include Test 2 scores.
- ○ No, I advise them not to include Test 2 scores.
- ○ I do not have any advice for them.

Feel free to elaborate on your advice below. Those who see your advice will view this explanation as well.

Attitudes towards test-blind and optional policies

Thank you for your answer! You have finished all the admission decisions. Next, we will ask about your views on two issues in US higher education.

In recent years, some universities have adopted a test-blind admission policy. That is, they will not consider students' SAT and ACT scores when they make admission decisions. Between this policy and the traditional test-required policy under which all applicants must submit their SAT or ACT scores, which one do you support?

- ○ I strongly support the test-blind policy
- ○ I somewhat support the test-blind policy
- ○ I am indifferent
- ○ I somewhat support the test-required policy
- ○ I strongly support the test-required policy

Can you briefly explain your reasoning on this?

In recent years, many universities have adopted a test-optional admission policy. That is, they will not require students to submit their SAT or ACT scores in their college applications, but may still consider the scores if they are submitted. Between this policy and the traditional test-required policy under which all applicants must submit their SAT or ACT scores, which one do you support?

- ○ I strongly support the test-optional policy
- ○ I somewhat support the test-optional policy
- ○ I am indifferent
- ○ I somewhat support the test-required policy
- ○ I strongly support the test-required policy

Can you briefly explain your reasoning on this?

Thank you for your answers. You have completed the main part of the survey. At the end, we would like to know more about you.

Demographics

How old are you?

- ○ Under 18
- ○ 18-24 years old
- ○ 25-34 years old
- ○ 35-44 years old
- ○ 45-54 years old
- ○ 55-64 years old
- ○ 65+ years old
- ○ Prefer not to say

How do you describe yourself?

- ○ Male
- ○ Female
- ○ Non-binary / third gender
- ○ _____ Prefer to self-describe
- ○ Prefer not to say

Choose one or more races that you consider yourself to be

- ☐ White or Caucasian
- ☐ Black or African American
- ☐ American Indian/Native American or Alaska Native
- ☐ Asian
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ Other
- ☐ Prefer not to say

What is the highest level of education you have completed?

- ○ Some high school or less
- ○ High school diploma or GED
- ○ Some college, but no degree

- ○ Associates or technical degree
- ○ Bachelor's degree
- ○ Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)
- ○ Prefer not to say

What best describes your employment status over the last three months?

- ○ Working full-time
- ○ Working part-time
- ○ Unemployed and looking for work
- ○ A homemaker or stay-at-home parent
- ○ Student
- ○ Retired
- ○ Other
- ○ Prefer not to say

What was your total household income before taxes during the past 12 months?

- ○ Less than $25,000
- ○ $25,000-$49,999
- ○ $50,000-$74,999
- ○ $75,000-$99,999
- ○ $100,000-$149,999

- ○ $150,000 or more
- ○ Prefer not to say

How do you describe your political leaning?

- ○ Very liberal
- ○ Somewhat liberal
- ○ Neither liberal nor conservative
- ○ Somewhat conservative
- ○ Very conservative
- ○ [ ] Other (please specify)
- ○ Prefer not to say

Comments

If you have any suggestions or comments on this survey, especially if you find any part confusing, please write them below. Then please click Next and you will be redirected back to Prolific. You will receive your payment in two days. Thank you for taking our survey!