Why Exclude Test Scores from Admission Decisions? *

Yucheng Liang† Wenzhuo Xu‡

[Preliminary and Incomplete]

Latest version will be updated here

April 30, 2024

Abstract

One major argument in support of test-optional and test-blind college admission policies is that standardized test scores inaccurately reflect students' abilities and are biased against those with fewer resources. This argument goes against standard economic reasoning as information, even if noisy or biased, never has negative value. In an experiment, we show that participants responsible for admitting students for an educational opportunity are indeed willing to exclude invalid or biased test scores from their admission criteria. This result is primarily driven by procedural fairness concerns and an underestimation of the scores' usefulness. However, this underestimation can be mitigated through experience in making admission decisions both with and without these test scores.

*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 allows 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 has recently reignited as several highly selective colleges have reinstated their SAT/ACT requirements.[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. 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. 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 (Chetty et al., 2023). Given that colleges have the freedom to use test

[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/

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 in test scores lead spectators to exclude them from the admissions process and explore the rationales behind such exclusions.

In the main experiment, each spectator is tasked with selecting students from an introductory data science 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. A 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, 82% of spectators prefer 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 influenced by biased, and to a lesser extent, invalid 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 reduces 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, given the same decision weight on Test 2 scores in admission decisions, spectators are more likely to prefer excluding the scores when the test prep is biased. This suggests that the moral cost arising from concerns over procedural fairness influences the preference to exclude

Test 2 scores. 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. Third, 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.

Although many spectators justify the exclusion of Test 2 scores on the grounds that they are not very useful, our findings indicate that this perception is not accurate. Even among spectators who say the scores are useless, 60% use the scores in at least one admission decision. To examine whether spectators update their perceptions of score usefulness based on their experience of making admission decisions, we elicit their 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 increase across the board. These results imply that the exclusion of invalid or biased test scores is partly driven by an initial underestimation of their usefulness, which can be corrected through the experiential learning process inherent in making admission decisions. They also resonate with recent findings that the difficulty of thinking through possible realizations of information can lead to suboptimal information demand (Liang, 2023).

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, exclusion of Test 2 scores is not significantly correlated with response time for admission decisions, which is inconsistent with the anticipated decision cost hypothesis. Third, when asked to provide advice to other spectators on whether to include or exclude Test 2 scores, most spectators give advice that is consistent 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 are applicable 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 for applicant background information such as race. 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.

Despite its tightly controlled approach, a potential limitation of the experimental paradigm is its exclusion of potentially significant real-world motivations for test-optional and test-blind policies, such as strategic and general equilibrium considerations.[5] To determine if validity and bias constitute the primary considerations shaping attitudes towards test-optional and test-blind policies, 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 validity and bias of standardized tests are indeed the main considerations. Moreover, the support for these policies is strongly correlated with the decision to exclude biased Test 2 scores in the experiment. These results demonstrate that the preferences identified in our controlled experiments align closely with 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 their emphasis on standardized tests' invalidity and

[5]For example, these policies may help colleges alleviate social pressure and fend off legal challenges on their admission criteria. They may also reduce the financial and mental costs that students incur on test preparation.

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. Several papers have proposed alternative motivations, including strategic and general equilibrium factors, for omitting test score requirements.[6] Among these studies, Dessein et al. (2023) is closest to ours. They contend 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, it specifically investigates perspectives on admission policies, particularly the inclusion of test scores in admission criteria, rather than focusing on the resultant admission outcomes.

By studying information preferences for admission decisions, we connect two behavioral economics literatures. One literature studies distribution decisions when recipients' types are uncertain (Cappelen et al., 2022, 2023, 2024; Chakraborty and Henkel, 2024). Although our findings on admission decisions contribute to this body of evidence, our research primarily focuses on information preference, whereas these studies only 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, thus enabling us to isolate the influence of social preferences on information choice. Moreover, these studies do not address information avoidance (see Golman et al. (2017) for a review) which is our focus. 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 dis-

[6]For instance, Conlin et al. (2013) argue that test-optional policies might enhance the average *submitted* SAT scores of their enrolled students, potentially boosting their rankings.

parities in confidence reporting between men and women with equivalent performance levels. 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. 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 of 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 (Bartling et al., 2014) and opportunities (Akbaş et al., 2019). 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 drives the exclusion of invalid and biased test scores. In computer science, a large and fast-growing literature on algorithmic fairness studies how to select and process information to design fair allocation algorithms. Oftentimes, this entails the exclusion of useful but biased information, a practice Rambachan et al. (2020) criticizes for its inconsistency with consequentialist principles. Our findings provide a potential explanation for why information exclusion is a popular approach to achieving algorithmic fairness.

Our paper is also related to the literature on bounded rationality in the demand for useful information (Ambuehl and Li, 2018; Charness et al., 2021; Guan et al., 2023; Liang, 2023; Guan, 2023). A common finding of this literature is that frequently misjudge the usefulness of information. Specifically, Liang (2023) links this failure to the difficulty of processing various potential outcomes of information and synthesizing the payoffs from these contingencies. We find that spectators' demand for test scores increases after they accumulate admission-related experience.

This could be explained by the difficulty of contingent reasoning as admission experience could improve spectators' understanding of scenarios where test scores are useful.

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 science course. Upon finishing the course, these students took two tests, referred to as Test 1 and Test 2, covering different course content areas. Each exam consisted of five questions with a total possible score of 10. For Test 1, students did not receive any additional test preparation. 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 one 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 select students from the introductory course to enroll in an advanced data science course. Each spectator makes 7 rounds of admission decisions, each time selecting three out of eight students. They could also choose an option that says "I cannot decide which 3 students to admit." The 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.

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 first scenario is a "No Prep" condition where no student received any additional preparation for Test 2. The second scenario is randomized among

spectators and includes one of the following conditions:

- **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.

We randomize the order in which spectators encounter these two scenarios.

After reporting their information preferences for the two scenarios, each spectator gets to know which one is the actual test prep situation for the students considered for admission. For this scenario, we ask spectators to confirm their previously stated information preference by completing a small real effort task, which entails typing in a sentence. A confirmation would ensure that the observability of Test 2 scores adheres to their stated preference in the majority of admission rounds.

Spectators know that one of the seven rounds consists of real students who took our introductory course while the other six rounds are fictitious. Unbeknownst to them, the real round is the last one, which is also the only one where the observability of Test 2 scores is affected by their reported information preference. Whether Test 2 scores are revealed is fixed for the fictitious rounds: they are in Rounds 4 to 6 but not in Rounds 1 to 3.

Table 1 lists the students' test scores and family income status for the six fictitious rounds under the No Prep scenario.[7] These numbers are specifically designed. The first three rounds allow us to identify spectators' social preferences such as preference for meritocracy and lower-income students. Rounds 4 and 5 allow us to observe spectators' tradeoffs between the two test scores, while Round 6 identifies the tradeoff between high Test 2 scores and lower-income status. The orders within the first and second three rounds are randomized.

After the sixth round of admission decision, we elicit spectators' information preference for a second time. This elicitation is intended to measure if experience with admission decisions

[7]For the other three test prep scenarios, the only difference is that the Test 2 scores for students who received test prep are set to be one point higher.

Table 1: Student information in the six fictitious admission rounds

Round	Higher-income students	Lower-income students
1	(9,-), (8,-), (5,-), (4,-)	(9,-), (7,-), (6,-), (4,-)
2	(8,-), (8,-), (7,-), (0,-)	(6,-), (5,-), (5,-), (5,-)
3	(7,-), (7,-), (7,-), (6,-)	(7,-), (7,-), (7,-), (7,-)
4	(9,9), (6,5), (5,5), (5,5)	(8,8), (7,6), (6,8), (5,5)
5	(9,8), (8,6), (7,8), (3,4)	(9,9), (6,5), (6,5), (6,5)
6	(7,9), (7,8), (7,5), (6,5)	(7,9), (7,7), (7,5), (7,4)

Notes: The numbers in each parenthesis represent a student's scores in Test 1 and Test 2. Test 2 scores are revealed in Rounds 4 to 6 but not in Rounds 1 to 3.

affects information preferences. Right after this elicitation, we also ask spectators to advise other participants who are in the same test prep scenario on whether to request access of Test 2 scores. The advice reflects spectators' information preferences when others decide whom to admit.

The last round of admission decision is made after the advice elicitation, where the disclosure of Test 2 scores is based on a random selection of the spectators' expressed information preferences at one of the two junctures. Finally, we survey participants on their attitudes towards test-blind and test-optional admissions policies and collect demographic information.

Incentives. The elicitations of information preferences and admission decisions are all incentivized. Spectators are told that their admission decisions for the round that consists of real students have a chance of being implemented. For the first elicitation of information preferences, after the true test prep scenario is revealed, we ask spectators to confirm their preferences for this scenario by typing in a sentence unless they reported indifference. They are told that the confirmed preference will affect whether the majority of the admission rounds have Test 2 scores available or not. If they do not type in the sentence, they are told that there may be more rounds where the observability of Test 2 scores does not adhere to their reported preferences. For the second elicitation of information preferences after the sixth round of admission decisions, spectators are told that one of their two reports of information preferences will determine whether Test 2 scores

are revealed for the last admission decision. The advice to other spectators has a chance of being sent out for real.

Logistics. We recruit 900 spectators from Prolific on December 18, 2023, and 596 of whom participate in our main treatment. Each participant receives a fixed payment of $5 and the median time spent is 17.5 minutes. We conduct several additional robustness and mechanism treatments, one of which (Performance Prediction) includes an additional $3 incentive bonus. These additional treatments will be introduced later in the paper.

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 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' family income status, ability, and potential. The admission decision is denoted by a which is a vector of ones and zeros representing whether each student is admitted or rejected. Prior to decision, the spectator decides whether to have access to students' test scores T with the probability of its realizations denoted by $p(t)$. All the other information she observes about the students is summarized by S, which will be suppressed below for ease of notation.

The spectator's preference for admission outcomes is described by the value function $v(a, \theta)$. Note that this function is general enough to accommodate social preferences such as meritocracy, preference for low-income students, preference for diversity, etc.

3.2 Standard consequentialism

A standard consequentialist spectator maximizes the expected value of admission outcomes given available information. Therefore, the value of test scores is

$$V(T) = \sum_t p(t) \cdot E[v(a(t), \theta) - v(a(\emptyset), \theta)|t], \tag{1}$$

where $a(t)$ is the optimal admission decision given test scores t, and $a(\emptyset)$ is the optimal decision when t is not observed. Because for each t, $a(t)$ yields a weakly higher value than $a(\emptyset)$, the value of test scores is always nonnegative. Hence, the spectator never strictly prefers to exclude T.

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 T then becomes

$$V(T) = \sum_t p(t) \cdot E[v(a(t), \theta) - v(a(\emptyset), \theta)|t] - C(T). \tag{2}$$

With procedural fairness concerns, whether to include or exclude the test scores depends on the tradeoff between the expected usefulness of the scores and the moral cost of including them. This leads to an important comparative statics: test score exclusion increases as the expected usefulness of the scores goes down.

3.4 Decision costs

Another potential reason for test score exclusion is that admission decisions with the test scores present may be more costly than those without. It could be the mental cost of sifting through and

digesting additional information. It could also be the increased social pressure if the test scores are also observed by third parties. Specifically, with the test scores present, the spectator chooses $\tilde{a}(t)$ to maximize $E[v(a,\theta)|t] - c(a,t)$. The value of test scores is then

$$V(T) = \sum_t p(t) \cdot E[v(\tilde{a}(t),\theta) - c(\tilde{a}(t),t) - v(a(\emptyset),\theta)|t]. \tag{3}$$

One implication of decision costs as a potential driver of test score exclusion is that spectators would be less willing to exclude the test scores if other people make the admission decisions and, hence, bear the decision costs.

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 the test scores, she may be willing to blind herself from them. Formally, the spectator anticipates that when she observes the test scores t, she will choose $\hat{a}$ to maximize $E[\hat{v}(a,\theta)|t]$ where $\hat{v}$ is a biased value function different from v. Therefore, the value of test scores,

$$V(T) = \sum_t p(t) \cdot E[v(\hat{a}(t),\theta) - v(\hat{a}(\emptyset),\theta)|t], \tag{4}$$

could be negative.

This explanation of test score exclusion also has implications for when admission decisions are made by other people. Assuming that other people's decisions are more likely to be misaligned than one's own, people should be more willing to exclude test scores for other people's admission decisions.

4 Results on Admission Decisions

We begin by analyzing the aggregate patterns of spectators' admission decisions for the six factitious groups. These decisions inform us on 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. 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 sometimes lower when they are affected by test prep that is unavailable to lower-income students.

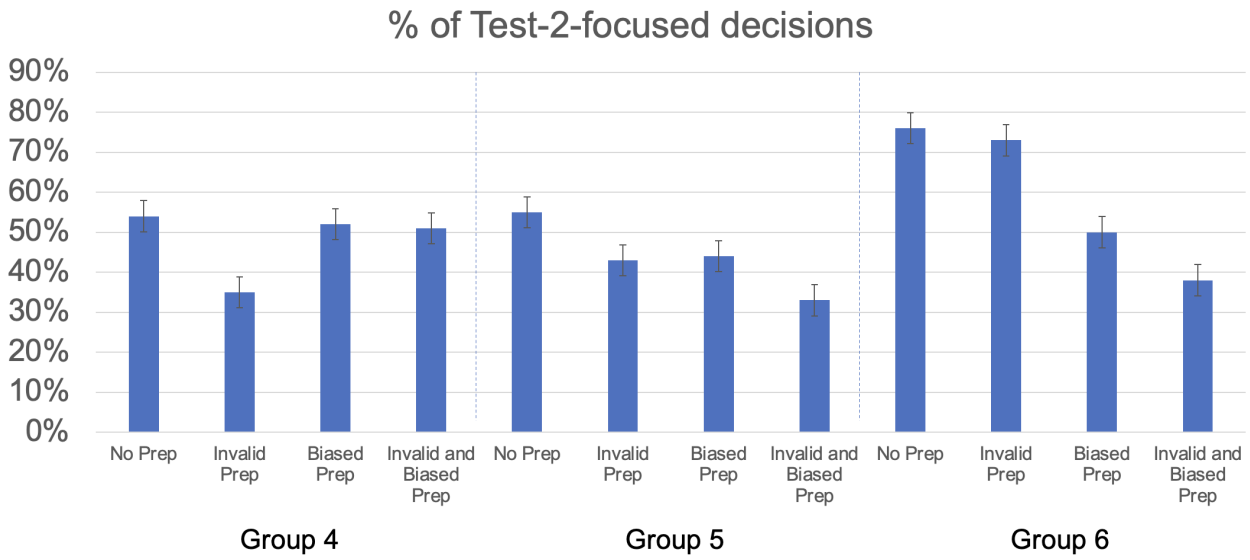

Figure 1: Proportions of Test-2-focused admission decisions for Groups 4 to 6

Specifically, the student information in these groups is designed such that each group includes one higher-income and one lower-income student whose scores in both tests dominate all other students. 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 (94% for Group 4, 91% for Group 5, and 87% 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 1 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 an higher score in 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 vs. Invalid Prep). 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 vs. Invalid and Biased Prep). 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 of them is a higher-income student with a higher Test 2 score and the other is a lower-income student. Again, both the invalidity and the bias of the test prep lead to fewer Test-2-focused decisions.

5 Information Preferences under Different Test Prep Scenarios

Do spectators prefer that Test 2 scores be revealed when they make admission decisions? Figure 2 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 preference to exclude Test 2 scores is mostly strict: 92% of these participants, when prompted, complete the real effort task to confirm their exclusion preference. This result indicates that the invalidity and bias of a test make people more willing to exclude it from the admission process.

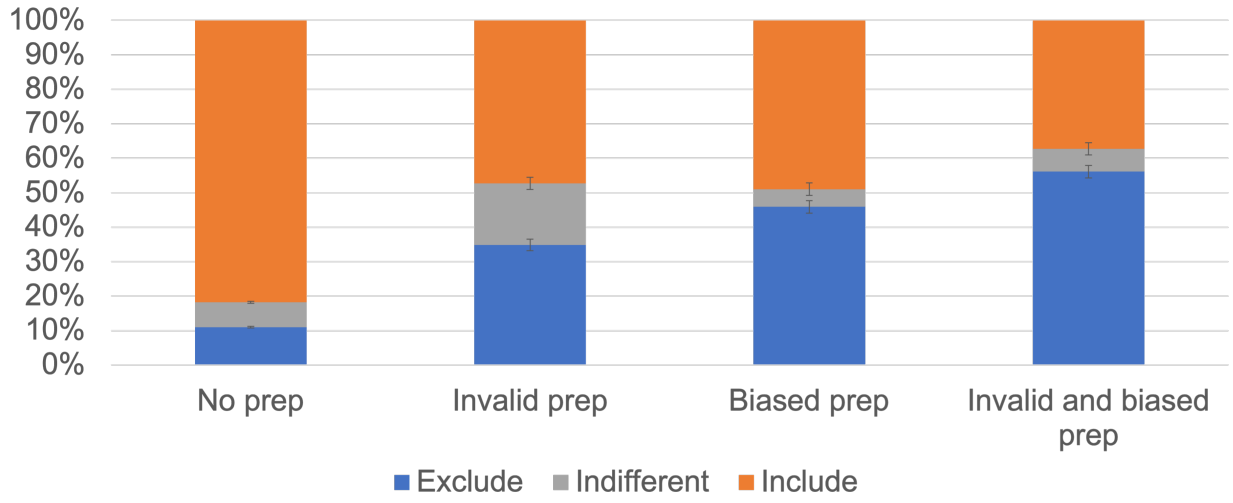

Figure 2: Information preferences across test prep scenarios (1st elicitation)

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 4, a large language model. The exclusion of Test 2 scores predominantly revolves around two considerations – the fairness and usefulness of the test scores. We then hand-code each answer as whether it mentions these two considerations. In the two scenarios where only higher-income students receive test prep, 86% (Biased) and 74% (Invalid and Biased) of spectators who prefer to exclude Test 2 scores mention that including the scores would be unfair. This number drops to 37% in the Invalid Prep scenario and further to 11% in the No Prep scenario. In the three scenarios with test prep, 69% (Invalid), 73% (Biased) and 66% (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 52%. In addition to "unfair" and "not useful," we also look for answers that mention decision costs and anticipated 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 support the procedural fairness model illustrated 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. As a result, perceptions of unfairness and low usefulness are two potential motivations for excluding test scores.

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 science 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. 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 ($p < 0.001$). This implies that fairness concerns are an important driver for the preference to exclude invalid and biased test scores from admission decisions.

One may wonder whether test score exclusion is driven by spectators' fairness concerns about the admission procedures or 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, preferences to include or exclude Test 2 scores for any scenario with

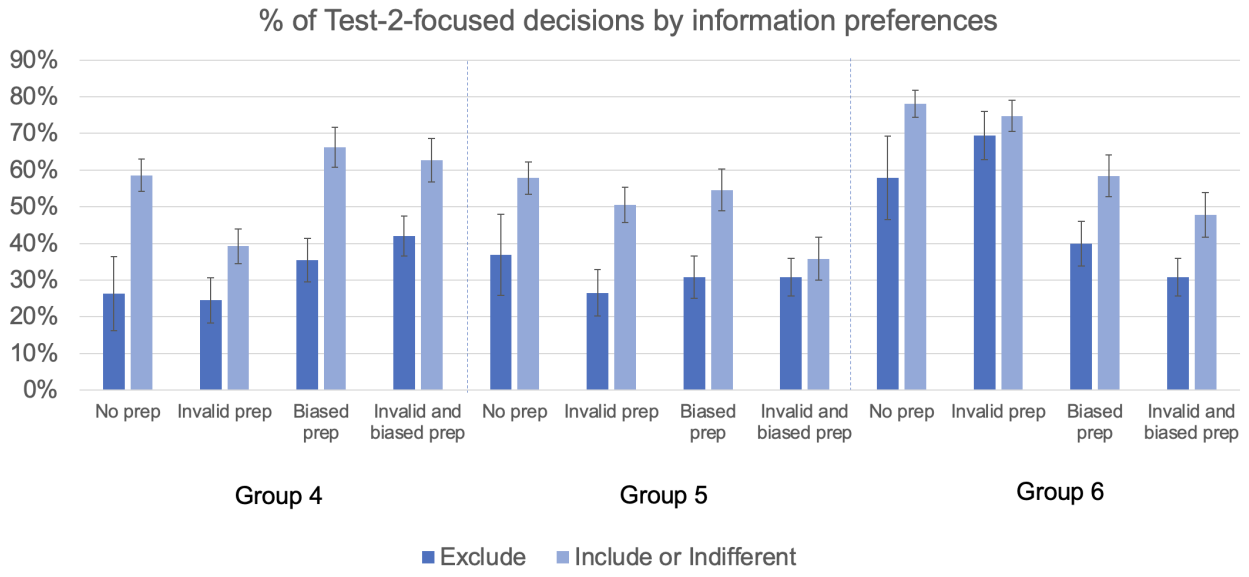

Figure 3: Proportions of Test-2-focused admission decisions by information preferences

test prep are uncorrelated with two measures of outcome-based social preferences, the number of meritocratic decisions and the number of admitted lower-income 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 perceptions of test score usefulness, 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 3 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. This result is consistent with the procedural fairness model. Conversely, it is not consistent with the anticipated decision mistakes model.

6.4 Underestimation of Score Usefulness

While spectators who prefer to exclude Test 2 scores make fewer Test-2-focused decisions, it remains unclear whether their use of Test 2 scores was correctly anticipated when they reported

their information preference. This question can be addressed by comparing information preferences before and after spectators make 6 rounds of admission decisions. This is because if spectators initially misperceived the usefulness of Test 2 scores, their subsequent admission experience could correct these misperceptions, which in turn would affect their demand for the scores. As shown in Figure **??**, demand for Test 2 scores increases across all scenarios, particularly in those involving test preparation, after the admission rounds. Furthermore, this increase in demand correlates strongly with the number of Test-2-focused decisions in Groups 4 to 6. These findings suggest that spectators may initially underestimate the usefulness of Test 2 scores when opting to exclude them, but this underestimation tends to decrease with experience in making admission decisions.

This experimental result mirrors 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 instance, Caltech stated that "standardized testing provides admissions officers and faculty reviewers useful information about academic preparedness." 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.[8]

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. Moreover, the mistake-avoiding motive is inconsistent with the finding that the use of Test 2 scores in admission decisions is negatively correlated with preference for their exclusion (see Section **??**). We will now present more behavioral evidence that is inconsistent with these motivations.

[8] 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.

6.5.1 Decision time

We focus on the time a spectator takes to submit her admission decision as a measure of decision costs.[9] As Table **??** 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 insignificant if we consider the difference in decision time between decisions with and without Test 2 scores. This result indicates that decision costs are not a main driver of information preference.

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 preference can test the two alternative motivations. If a spectator prefers to exclude Test 2 scores in order to lower the costs of making admission decisions, she should be less likely to advise others to exclude the scores as the costs are not born 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.

Table **??** cross-tabulates the advice against spectators' own information preference elicited right before the advice. First and foremost, there are few spectators who give advice opposite to their own preference. Among the four test prep scenarios, only the No Prep scenario has a substantial proportion of spectators who prefer to exclude Test 2 scores for themselves advising others to include them. This suggests that decision costs are not a main reason for excluding Test 2 scores when they are invalid or biased. For spectators who prefer inclusion for themselves, advice in the opposite direction is even rarer, suggesting that anticipated decision mistakes are unlikely to be a pivotal factor for information preference. A number of spectators who are either in favor of or

[9] An alternative measure is decision avoidance, i.e., selecting "I cannot choose whom 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.

against including Test 2 scores themselves decide not to provide any advice to others. Most of this type of behavior is justified by a 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 the external validity of our findings.

6.6 Other Applicant 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. [10] Also, after the Supreme Court's decision that bans the use of race in college admissions, people predict that more colleges will stop requiring standardized test scores because the value of these scores are diminished when colleges are no longer able to adjust for the racial gap.[11]

To investigate the effects of other applicant information on preferences to include or exclude

[10] 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).

[11] https://www.forbes.com/sites/vinaybhaskara/2023/07/05/how-the-end-of-affirmative-action-will-impact-college-admissions/?sh=bc5dfea32db7. See also **?**.

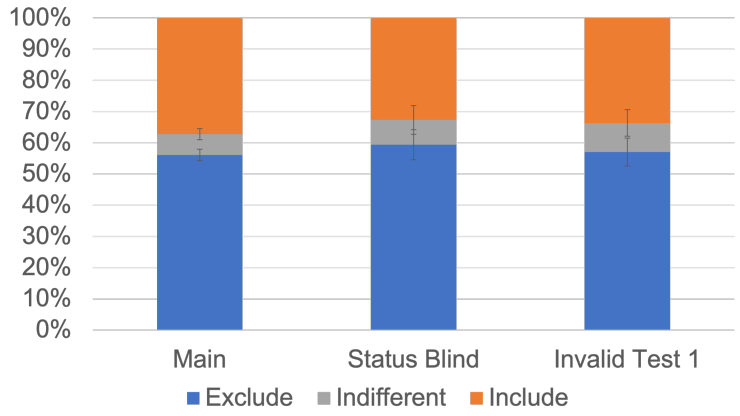

Figure 4: Information preferences across experiments: Invalid and Biased Prep

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 go down when spectators cannot use other applicant information to adjust for the scores' bias.

Figure 4 shows that the demand for Test 2 scores in these two treatments are the same as the main treatment. This result indicates that although the value of test scores should depend on other available applicant information, people tend to evaluate the scores in isolation when they consider their demand.

References

M. Akbaş, D. Ariely, and S. Yuksel. When is inequality fair? an experiment on the effect of procedural justice and agency. *Journal of Economic Behavior & Organization*, 161:114–127, 2019.

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

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.

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.

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.

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.

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.

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.

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

[TBA]