Excluding Flawed but Useful Information for Fair Allocation Decisions*

Yucheng Liang† Wenzhuo Xu‡

October 30, 2025

Abstract

When allocating scarce resources, decision makers must determine what information about potential recipients to include in their allocation criteria. In a series of experiments, we find that participants making student admission and bonus allocation decisions frequently exclude information that contains biases or noise. This exclusion reflects a tradeoff between fairness concerns and the usefulness of information. These fairness concerns are primarily focused on the allocation procedure rather than the resulting outcomes.

*This study is approved by CMU IRB in Protocols 2023_00000333 and 2023_00000337. The AEA RCT registry IDs are AEARCTR-0012709 and AEARCTR-0016492. Kate Yixin Huang, Jack Tianrui Lin, and Allison Tribendis provided excellent research assistance. Julie Downs, Christine Exley, Sean Fath, Luca Henkel, David Huffman, Navin Kartik, David Munguia Gomez, Muriel Niederle, Jason Somerville, Lise Vesterlund, Erina Ytsma, Sevgi Yuksel, and many other seminar and conference participants offered valuable comments. AJ Skiera kindly provided the raw data for a public opinion poll on standardized tests. This paper subsumes and extends an earlier draft titled "Why Exclude Test Scores from Admission Criteria?" Errors are ours.

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

‡The Hong Kong Polytechnic University. Email: wen-zhuo.xu@polyu.edu.hk.

1 Introduction

When allocating scarce resources, decision makers must determine what information about potential recipients to use as allocation criteria. A growing concern is that including biased or invalid information—signals that contain systematic bias or noise—may lead to unfair decisions. For example, judges consider past arrest records when making pretrial release decisions. While such records may indicate a defendant's flight risk or threat to public safety, they can also reflect discriminatory policing practices (Pierson et al., 2020). In hiring, referrals and recommendation letters may signal candidate quality, but minority applicants often face disadvantages in securing them due to limited access to professional networks (Hoffman, 2017). Standardized test scores like the SAT and ACT predict college success (Chetty et al., 2023), but access to costly test preparation can undermine their validity and introduce systematic bias against students from disadvantaged backgrounds (Buchmann et al., 2010).

There are two broad approaches to addressing concerns about biased or invalid information. One is to include such information in allocation criteria and design decision rules that extract useful signals while minimizing unfair outcomes. The other is to exclude the information entirely from consideration. While the former, more nuanced approach is usually favored by economists and computer scientists (Rambachan et al., 2020), the latter, more categorical approach is more often reflected in popular policy proposals. For example, in response to concerns about biased data inputs, some scholars and advocacy groups have argued that algorithmic risk scores "have no place in pretrial justice" (Minow et al., 2019). In college admissions, many universities have adopted test-optional or test-blind policies, allowing or even requiring applicants to omit standardized test scores, largely on the grounds that these scores are invalid and biased (Feder and Bello, 2024).

While concerns about the bias and validity of decision criteria may be well-founded, it remains unclear why such concerns are seen as good justification for policies that voluntarily discard information.[1] From a standard economic perspective, more information is always (weakly) beneficial

[1] In the context of college admissions, Dessein et al. (2025) refer to this as the "test-optional puzzle." They show that, under broad conditions, test-blind and test-optional policies constitute a form of information avoidance.

for achieving the decision maker's objective — even when that objective incorporates fairness in outcomes. What kinds of concerns, then, lead people to support policies that deliberately forgo potentially useful information?

In this paper, we study information preferences in a series of experiments where participants ("spectators") allocate scarce resources. Our study addresses two main research questions:

1. Are spectators more likely to exclude information from allocation criteria when it is invalid or biased?

2. Are such exclusion decisions driven by fairness concerns — and if so, what kind?

We use the experimental method to explore these research questions because it allows us to vary the validity and bias of information, as well as the fairness relevance of the decision, in a controlled way — something that is difficult to achieve in observational settings. The experimental design also enables us to observe both participants' information choices and their allocation decisions, allowing us to analyze the relationship between the two. This, in turn, helps us identify the motivations underlying information exclusion. We recruit participants from Prolific, an online survey platform, to capture the views of the general public. Public attitudes are particularly relevant given their role in shaping social norms and institutions (Almås et al., 2020).

In the main experiment, each spectator selects students for admission to an advanced data analysis course and decides whether to include a potentially biased or invalid test score in the admission criteria. We use this decision context to evoke considerations relevant to the ongoing debate over the role of standardized tests in admissions. Specifically, the students under consideration come from different family income backgrounds and have all completed two tests—Test 1 and Test 2—that assess their basic data analysis skills. Spectators are informed that, in the absence of test preparation, students from higher- and lower-income backgrounds perform similarly on both tests. However, while no students received test prep for Test 1, some may have had access to 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 each of these conditions, test preparation may increase the recipient's Test 2 score by up to 2 points out of a total of 10. 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 changes in the broader information environment and to who makes the admission decisions.

After reporting their preferences for including or excluding Test 2 scores in two test prep scenarios, each spectator is randomly assigned one of the scenarios to be the true one they face, under which they make seven 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 reveal that most spectators prefer students with higher test scores, while some also show a preference for lower-income students. When Test 2 scores are available, the weight spectators place on them varies systematically with the test prep scenario. This decision weight is lower when the scores are affected by test prep that either does not enhance data analysis skills or is inaccessible to lower-income students.

Moreover, by cross-tabulating admission decisions and information preferences, we find a strong negative association between the use of Test 2 scores and the preference to exclude them. First, spectators who initially prefer to exclude Test 2 scores are less likely to rely on those scores in their subsequent admission decisions when the scores are revealed. Second, when we elicit spectators' inclusion or exclusion preferences a second time—after they complete six rounds of admissions—the overall preference for exclusion decreases across all scenarios involving test prep. This decline is driven primarily by spectators whose prior admission decisions heavily relied on Test 2 scores.

The finding that information exclusion increases when the information is invalid or biased, but decreases as spectators use it more, suggests that information preferences reflect a tradeoff between usefulness and some "cost" associated with invalidity and bias. To investigate whether these costs stem from fairness concerns, we conduct a treatment called Performance Prediction, in which we remove the social impact of spectators' decisions. Specifically, spectators see the same information—income status and test scores—but use it to predict students' performance in the advanced course, rather than to make admission decisions. These predictions do not affect the students in any way, and spectators are paid based on prediction accuracy. Compared to the main treatment, the proportion of spectators who choose to exclude invalid and biased test scores falls by half. This sharp reduction suggests that much of the exclusion behavior in the main treatment is driven by fairness concerns when decisions have consequences for others.

Why do fairness concerns lead to a preference for excluding biased or invalid information? Conceptually, concerns about the fairness of *outcomes*—that is, who receives which allocation—cannot by themselves explain information exclusion: in standard models, more information should always (weakly) improve the ability to achieve fairer outcomes, and decision makers are always free to ignore information they consider unhelpful. For outcome-based fairness concerns to generate exclusion, they must interact with bounded rationality—such that the presence of invalid or biased information either leads the decision maker to mistakenly choose less fair outcomes or increases the difficulty or discomfort of making allocation decisions. An alternative explanation is that fairness concerns are procedural: including biased or invalid information may undermine the integrity or legitimacy of the decision process itself, even if it does not affect the allocation outcomes or the cost of arriving at them. In this view, exclusion serves to preserve the fairness of the procedure, even at the expense of potentially useful information.

We find evidence in the admission experiment that concerns about outcome fairness are not the primary driver of information exclusion. However, to conclude definitively that exclusion is motivated by procedural fairness concerns, we must design an experiment that varies the validity and bias of information while holding constant both the distribution of allocation outcomes and the mental costs of making allocation decisions. This is not feasible in the admission experiment, as neither we nor the participants know the precise degrees of bias or invalidity introduced by test preparation. As a result, we cannot construct equivalent distributions over admission outcomes across different test prep scenarios. To address this limitation, we conduct a more stylized auxiliary experiment that allows for greater control over key elements of the design.

In the auxiliary experiment, spectators allocate a bonus between two *ex-ante* identical workers, Worker 1 and Worker 2, who have each completed two internet research tasks, Task 1 and Task 2. Task 1 scores accurately reflect both workers' performance, while the accuracy of Task 2 scores varies across two within-participant experimental parts. In the Accurate Info part, Task 2 scores are accurate for both workers. In the Invalid and Biased Info part, Worker 1's Task 2 score remains accurate, but Worker 2's score is inflated by 1 point with a 50% probability. This experimental design differs from the admission experiment in an important way: the data-generating process behind the bias and invalidity in Task 2 scores is fully specified. The context is also substantially different, allowing us to examine whether the preference to exclude invalid and biased information generalizes across settings.

As in the admission experiment, we are interested in spectators' preferences for including or excluding Task 2 scores from their allocation criteria. However, a key distinction is that in this new experiment, we elicit information preferences *after* spectators have already chosen and committed to a *full* contingent allocation plan—that is, specifying who receives the bonus under every possible realization of task scores, both when Task 2 scores are included and when they are excluded.

This ordering renders the mental costs of making allocation decisions sunk and thus irrelevant for subsequent information preferences. Moreover, because the allocation plan covers every possible contingency, we can control for the full distribution of allocation outcomes when examining how bias and invalidity affect information exclusion.

Despite the many changes in experimental design, the distribution of information preferences in the new experiment ends up remarkably similar to that of the admission experiment: 55% of spectators prefer to exclude Task 2 scores when they are invalid and biased, and 96% of these preferences are strict. Moreover, among those who strictly prefer exclusion, 82% had already committed to the same allocation plan that relies solely on Task 1 scores, regardless of whether Task 2 scores are included in the criteria. The fact that these spectators strictly prefer to exclude invalid and biased information even when it has no impact on the costs or outcomes of allocation decisions suggests that a substantial share of exclusion preferences is driven by non-consequentialist fairness concerns about the allocation procedure itself.

Beyond information preferences, spectators' choices of allocation plans also reflect the influence of non-consequentialist concerns. Even when a plan that uses invalid and biased Task 2 scores yields the same distribution of allocation outcomes as one based on accurate Task 2 scores, spectators are significantly less likely to choose the former. This suggests that non-consequentialist concerns not only motivate people to exclude biased or invalid information from the allocation criteria, but also reduce their willingness to act on such information when it is available.

Literature review. By studying information preferences for allocation decisions, this paper connects two behavioral economics literatures. One literature studies allocation decisions when recipients' attributes 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 allocation 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 allocation decisions that impact only the outcomes for others, not the evaluators' own payoffs. 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).[2] 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. 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 scores when making allocation decisions, many of them are still willing to give up access to this information.[3] Fath et al. (2022, 2023) show that some evaluators prefer to blind themselves to job applicants' race or gender to avoid making biased hiring decisions. Similarly, Saccardo and Serra-Garcia (2023) find that some advisors avoid learning about their own incentives to ensure their recommendations remain unbiased. Crucially, these studies focus on the avoidance of information that is irrelevant to normatively optimal decisions, whereas the information we study is instrumentally useful. We also consider bias avoidance as a potential motivation for information exclusion but find evidence against it in our settings.

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

[2] A large literature, following the seminal work of Dana et al. (2007), demonstrates that many individuals choose to remain ignorant of how their actions affect others, enabling them to make self-serving decisions without appearing selfish. Exley and Kessler (2023) extend these findings by showing that information avoidance can occur even in the absence of self-interest. While our research also examines information avoidance when self-interest is not at stake, we focus specifically on the validity and bias of information as key determinants in its exclusion from allocation criteria.

[3] Another difference between our study and Exley and Nielsen (2024) is that our spectators' allocation 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'.

procedures: the allocation of decision rights and opportunities (e.g., Bolton et al., 2005; Bartling et al., 2014; Akbaş et al., 2019). With allocation decisions becoming increasingly data-driven, information choice has become an important procedure in the decision-making process. Using a novel experimental design that holds the distribution of allocation outcomes constant while varying the validity and bias of information, our bonus allocation experiment demonstrates that procedural fairness concerns also drive the exclusion of invalid and biased information from allocation criteria. This finding also contributes to the literature on the demand for useful information (Ambuehl and Li, 2018; Charness et al., 2021; Liang, 2023; Guan, 2023; Guan et al., 2025) by offering a novel explanation for why demand may deviate from the instrumental value of information.

By setting our main experiment in the context of student admissions, we also contribute to the literature exploring the motivations 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.[4] Several papers have proposed alternative motivations, including strategic and general equilibrium factors, for omitting test score requirements.[5] Among these studies, Dessein et al. (forthcoming) is closest to ours. They propose that test-optional and test-blind policies can be used to manage 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*.

The rest of the paper is organized as follows. Section 2 describes the design of the admission experiment and Section 3 presents its main results. Section 4 examines the role of fairness concerns in the exclusion of biased or invalid information. Section 5 investigates the nature of the fairness concerns and the mechanisms behind their impact on information preferences. Section 6 concludes

[4]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.

[5]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.

and discusses the external validity and broader implications of our findings.

2 Admission Experiment: Design

Context. We conducted the experiment on Prolific using a US sample on December 18, 2023 with a participation fee of $5 (see screenshots in Appendix F).[6] At the outset, participants (henceforth "spectators") are informed about the context of the experiment, which is an introductory data analysis course we taught at a US university.[7] Students who completed this course took two tests, referred to as Test 1 and Test 2, which covered different course content areas.[8] Each test consisted of five questions, with each question worth 2 points, resulting in a maximum possible score of 10. For Test 1, students did not receive any additional test preparation beyond the course material, whereas for Test 2, some students might have received test prep.

Overview of admission decisions. Spectators are told that their main task in the experiment is to make admission decisions for the students who completed the introductory data analysis course and the two associated tests. Specifically, they will be presented with seven groups of students, each containing eight individuals. For each group, spectators are asked to select three students to be admitted to an advanced data analysis course. They are informed that one of the seven groups consists of real students from the introductory course, while the other six groups are fictitious. Their decisions for the real group have a chance of determining the actual admission outcomes for these students. Therefore, although spectators' admission decisions do not affect their own payments, they may carry real and meaningful consequences for the students. For any group, spectators may also choose an option labeled "I cannot decide which 3 students to admit," which forfeits their chance of determining the admission outcomes.

When making admission decisions, spectators will have access to each student's score from Test 1 and know whether the student came from a higher-income (self-reported family income $\geq$

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

[7] The courses in the experiment are not official university courses. They do not count toward students' GPAs and do not appear on their transcripts.

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

$100,000) or lower-income background (< $100,000). They may, in addition, also have access to the students' Test 2 scores. Spectators are informed that students from higher-income and lower-income backgrounds had similar performance in Test 1. The two groups also performed similarly in Test 2 when test prep was not provided.

Information preferences. After receiving an overview of the admission decisions, spectators are asked to indicate whether they prefer to have access to students' Test 2 scores when making these decisions. These information preferences are elicited under two different scenarios about test prep for Test 2, presented in random order:

- **No Prep Scenario**: No student received any additional preparation for Test 2.

- **Scenario with test prep**: Some students received additional test preparation designed to boost their performance on one question in Test 2, potentially increasing scores by up to two points. The exact nature and allocation of the test prep are randomized into one of the following conditions between spectators:

    1. **Invalid Prep**: All students received test preparation that could increase their Test 2 scores but did not enhance their data analysis skills. This is implemented by providing each student with the answer to one random question in Test 2.

    2. **Biased Prep**: Only students from higher-income backgrounds received test preparation that improves both their data analysis skills and their Test 2 scores. This preparation involved providing additional insights relevant to one test question.

    3. **Invalid and Biased Prep**: Only students from higher-income backgrounds received test preparation that could boost their Test 2 scores but did not enhance their data analysis skills.

These conditions create variations in the validity and bias of test scores because the nature and allocation of test prep directly affect the informativeness of Test 2 scores and the income gap in performance. Specifically, providing students with one free answer inflates scores without a corresponding skill improvement, making the scores noisier and less reflective of actual skills, thereby undermining their validity. In contrast, test prep that improves both data analysis skills and test scores ensures that the scores remain a valid measure of true abilities. When test prep is exclusively provided to higher-income students, it creates an additional advantage for them, resulting in Test 2 scores that are more biased in their favor.

In each scenario, spectators indicate their preference for accessing Test 2 scores when making admission decisions by selecting yes, no, or indifferent. The answers may have consequences, as one of the two scenarios represents the actual test prep situation for the students they will select from, and their reported preferences for this true scenario will affect whether Test 2 scores are revealed to them. In addition, spectators are asked to provide an open-ended justification for their answer in each scenario.

After reporting their information preferences, each spectator gets to know the true test prep scenario faced by the group of students they will admit from. 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.[9] 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.

Admission decisions: fictitious students. Spectators make admission decisions for seven groups of students under their true test prep scenario. They know that one of these groups consists of real students, but they don't know that the real group is the last one to be presented to them. It is also the only group for which the observability of Test 2 scores is affected by their reported information preferences. 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.

Figure 1 lists the test scores and family income status of the students in the six fictitious groups under the No Prep scenario.[10] The student information of these groups is specifically designed to

[9] 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.

[10] 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.

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.

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.[11] 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.

Spectators always make admission decisions for Groups 1 to 3 before Groups 4 to 6.[12] The orders within the first and second three groups are randomized. To make an admission decision for a group, spectators first choose how to order the eight 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

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

[12] We chose not to randomize the order between these two sets of groups for two reasons. First, admission decisions for Groups 1 to 3 are arguably easier as spectators need to digest less information, so completing these decisions first may help spectators ease into harder decisions later. Second, students in Group 3 and Group 6 have the same income and Test 1 score information. If Group 3 is presented after Group 6, some spectators may try to fill in the missing Test 2 scores with what they saw previously in Group 6, which complicates the analysis of Group 3's admission results.

to select exactly three students for admission unless they opt for the "I cannot decide" option.

Post-experience information preference and advice. 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.

Admission decisions: real students. The last admission decision is made after the advice elicitation. For each spectator, whether Test 2 scores are disclosed for this decision is equally likely to be determined by either her pre-experience or post-experience information preference.

Attitudes toward real world college admission policies. Finally, we survey spectators on their attitudes toward test-blind and test-optional admissions policies. In the first question, spectators are informed about the test-blind college admissions policy and asked whether they prefer it over the traditional test-required policy. In the second question, the test-blind policy is replaced with the test-optional policy, and spectators are asked the same question. For each question, spectators are also asked to provide an open-ended justification for their answer.

Other treatments. 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 $1 bonus for each correct prediction, 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, all of whom resided in the US, had completed at least 20 Prolific surveys, and maintained an approval rate of at least 99%. 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 median time spent on the survey was 17.5 minutes. Each spectator received a $5 participation fee, and those in the Performance Prediction treatment earned an additional accuracy bonus averaging $2.56.

3 Admission Experiment: Main Results

3.1 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 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, which also indicates that they care sufficiently about the admission decisions.[13] 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).

[13]94% of spectators who prefer to include Test 2 scores complete the real effort task to confirm their preference. Figure A1 shows the distributions of confirmed information preferences.

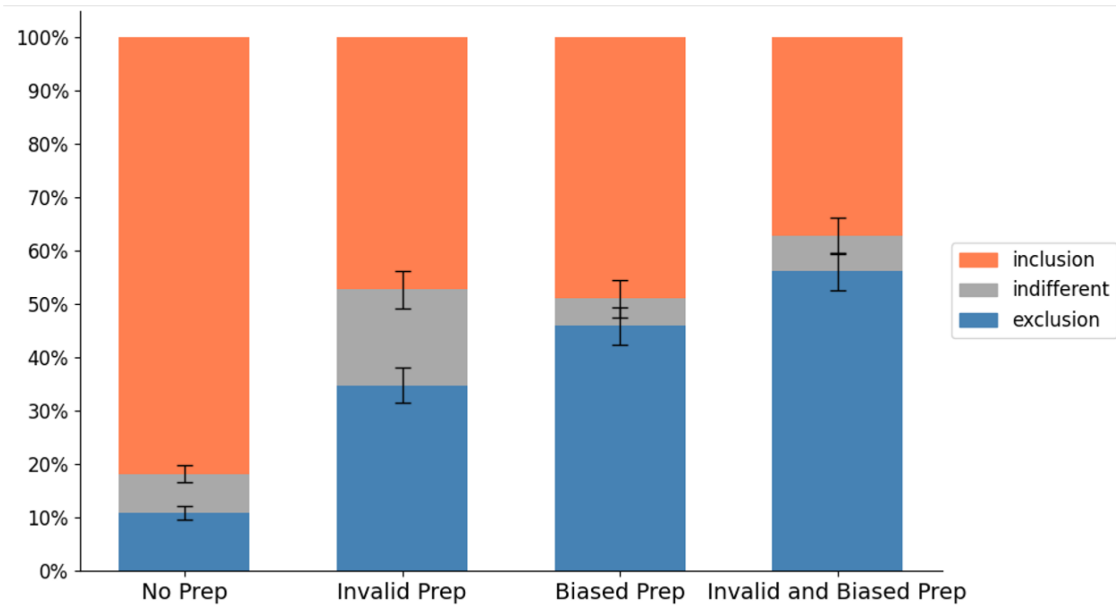


Figure 2: Information preferences across test prep scenarios (before admission experience)

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.

The preference to exclude invalid and biased test scores remains robust to the availability and quality of other information about the students. In the Invalid Test 1 treatment (N=107), Test 1 scores are randomly inflated but spectators don't know who the beneficiaries are. In the Status Blind treatment (N=101), spectators do not observe each student's family income background. Although these changes in the broader information environment could affect the value of Test 2 scores, in both treatments, the exclusion rates for invalid and biased Test 2 scores remain statistically indistinguishable from the main treatment (see Appendix B.1 for more detailed analysis and discussion of this result).

Information preferences are also largely invariant to who makes the admission decisions. When asked to advise others on whether they should request access to Test 2 scores, spectators' recommendations closely mirror their own information preferences when making admission decisions themselves (see Appendix B.2 for more detailed analysis and discussion of this result).

Taken together, the results indicate that invalidity and bias in test scores substantially increase people's willingness to exclude them from the admission process, and that this preference is robust across a range of informational and decision-making environments.

3.2 Admission Objectives and Information Usage

To understand spectators' information preferences, we must first understand their underlying admission objectives and how they use available information. These objectives and usage patterns are reflected in their admission decisions, which we analyze in this section.

3.2.1 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.[14] Almost everyone who does make a decision chooses three of the top performers.

3.2.2 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

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

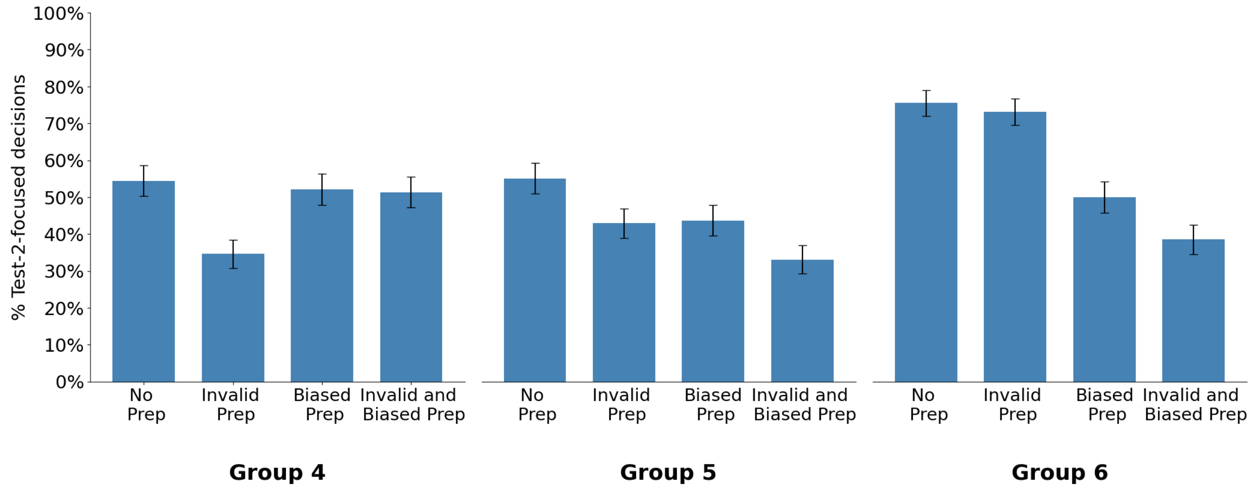


Figure 3: Proportions of Test-2-focused admission decisions for Groups 4 to 6

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.

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.

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 3 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 (Invalid Prep $-$ Invalid and Biased Prep $= 35\%$, $p < 0.001$). Moreover, holding the bias constant, spectators assign less weight to Test 2 scores when the test prep provided to higher-income students is invalid (Biased Prep $-$ Invalid and Biased Prep $= 11\%$, $p = 0.02$).

3.3 Relationship between Information Preferences and Information Usage

In most models, information preferences and information usage are tightly linked. Having analyzed them separately, we study their empirical relationship in this section.

First, spectators who initially prefer to exclude Test 2 scores make fewer Test-2-focused admission decisions when the scores are revealed. As Figure 4 shows, this relationship holds across all three fictitious student groups for which Test 2 scores are revealed and across all four test prep scenarios.

Second, the experience of making admission decisions increases the demand for biased or invalid Test 2 scores—particularly among spectators who rely heavily on those scores in their de-

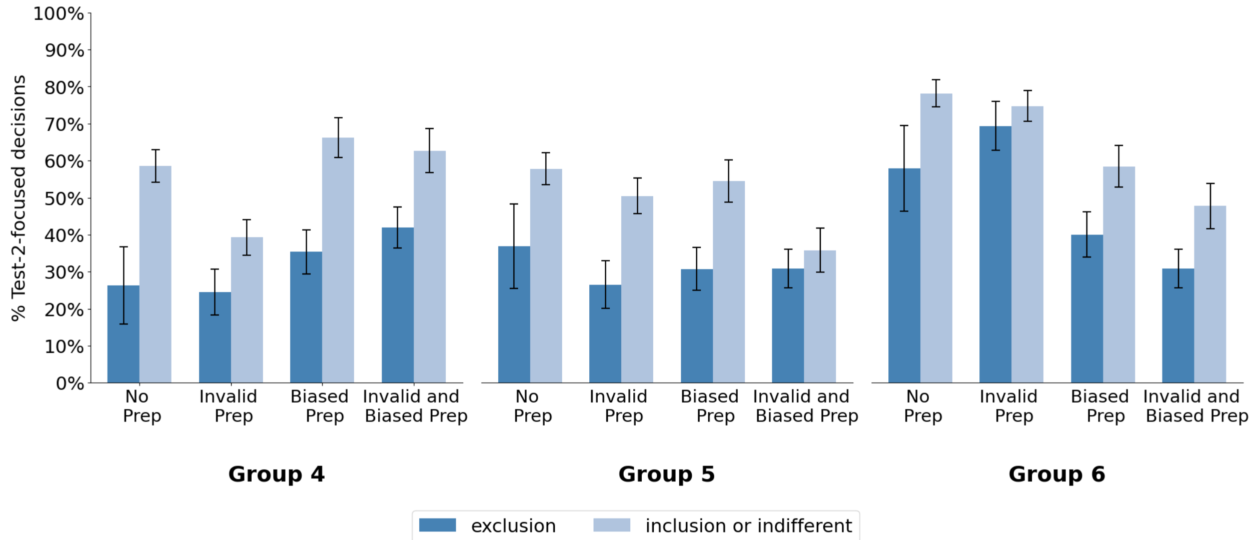


Figure 4: Proportions of Test-2-focused admission decisions by information preferences

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.

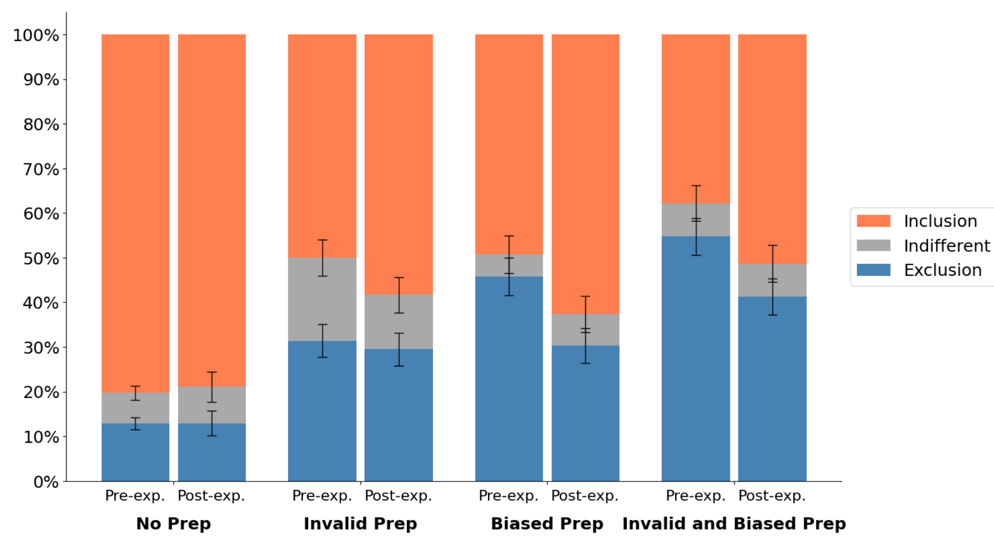


Figure 5: Information preferences across test prep scenarios (before and after admission experience)

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 1: 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. *** : $p < 0.001$; ** : $p < 0.01$; * : $p < 0.05$

cisions. We identify this experience effect by eliciting spectators' information preferences a second time, after they have completed six rounds of admission decisions. As shown in Figure 5, the demand for Test 2 scores increases across all three scenarios involving test preparation.[15] Moreover, Table 1 shows that this increase is especially pronounced among spectators who frequently make Test-2-focused decisions for Groups 4 through 6.[16] These results are consistent with a learning interpretation: as spectators make admission decisions, they come to recognize the instrumental value of Test 2 scores, and those who find them particularly useful become more likely to prefer their inclusion.

Taken together, these results indicate a strong negative association between the usage of biased or invalid test scores and the preference for their exclusion from admission criteria.

[15] Figure A3 further illustrates these changes with a transition matrix showing how individual preferences shift before and after admission experience.

[16] Ordered logit regression presented in Table A2 shows similar results.

4 The Role of Fairness Concerns in Information Exclusion

The finding that information exclusion increases when the information is invalid or biased, but decreases as spectators use it more, suggests that information preferences reflect a tradeoff between usefulness and some "costs" associated with invalidity and bias. To investigate whether these costs stem from fairness concerns, we conduct a treatment called Performance Prediction, in which we remove the social impact of spectators' decisions. 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–spectators are paid an additional $1 bonus for each correct prediction. 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.

Figure 6 shows the information preferences in this treatment in comparison with the main treatment. When tasked to predict student performance, 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.

The tradeoff between usefulness and fairness also emerges clearly from spectators' open-ended justifications for their information preferences in the main treatment. To systematically summarize these responses, we use GPT-3.5, a large language model, to identify the most frequently cited rationales. The summaries show that concerns about the usefulness and fairness of Test 2 scores are the two dominant reasons for preferring to exclude them.[17] We then hand-code each justification

[17]The primary rationale for including Test 2 scores is their perceived usefulness.

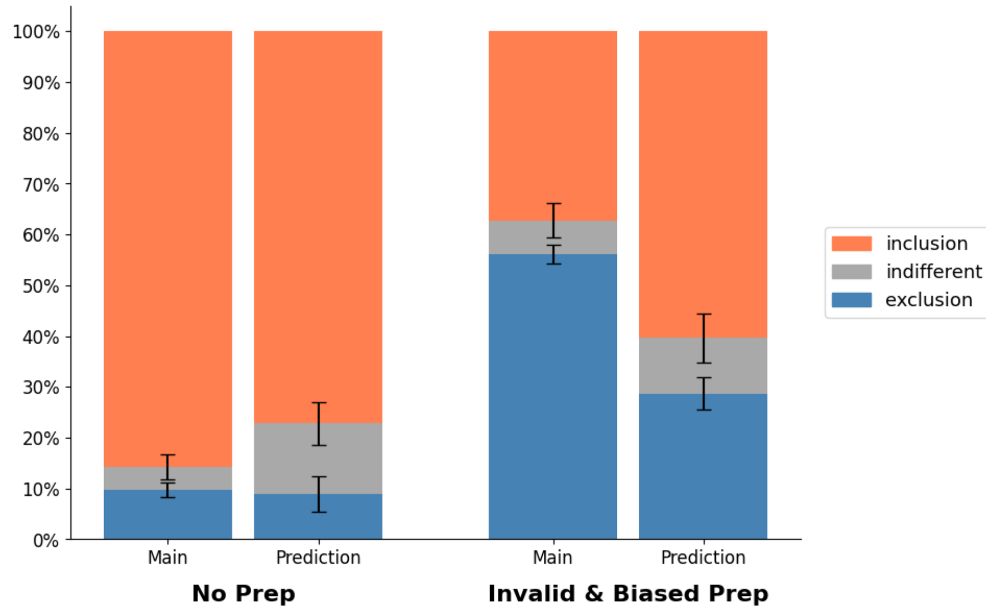


Figure 6: Information preferences in the Performance Prediction treatment

Notes: This figure shows the proportions of spectators who prefer to include, exclude, or are indifferent about the inclusion of Test 2 scores in the Performance Prediction treatment in comparison with the main treatment. Error bars represent the standard errors of the proportions of inclusion and exclusion preferences.

to identify the considerations mentioned, confirming that usefulness and fairness remain the two most commonly cited themes (see Appendix D for details on the summarization procedure, coding scheme, and additional results).

5 Outcome-Based vs. Procedural Fairness Concerns

Why do fairness concerns lead to a preference for excluding biased or invalid information? To address this question, we propose a conceptual framework that distinguishes between outcome-based and procedural fairness concerns, and we outline the channels through which each type of concern could give rise to information exclusion. We then present suggestive evidence from the admission experiment that is inconsistent with the outcome-based channels. Finally, to identify the procedural fairness channel more directly, we turn to an auxiliary experiment that elicits information preferences while holding constant the allocation outcomes associated with each information condition.

5.1 Conceptual Framework

5.1.1 Set-up

Consider an allocation problem with N candidates. An N-dimensional vector $\mathbf{a}$ specifies the allocation each candidate receives. Each candidate also has attributes relevant for allocation decisions—for instance, in an admission setting, these attributes may include family income, ability, and potential. Let $\boldsymbol{\theta}$ denote the vector of candidates' allocation-relevant attributes. While $\boldsymbol{\theta}$ is not directly observed, the spectator may observe a set of signals about it, such as family income categories or test scores.

The spectator first chooses which signals to include in the set of allocation criteria AC, with $\mathcal{S}(AC)$ denoting the set of possible realizations of the included signals. Based on these criteria, she then specifies an allocation plan $\pi(\cdot|\mathbf{s})$ that maps each realized vector of included signals $\mathbf{s}$ to a probability distribution over allocations $\mathbf{a}$.[18] Given the joint distribution of allocation-relevant attributes and signals $p(\boldsymbol{\theta}, \mathbf{s})$, the allocation plan π induces a joint distribution of attributes and allocations, which we refer to as the *outcome distribution*:

$$\mu^{\pi}(\boldsymbol{\theta}, \mathbf{a}) = \sum_{\mathbf{s} \in \mathcal{S}(AC)} p(\boldsymbol{\theta}, \mathbf{s}) \pi(\mathbf{a}|\mathbf{s}). \tag{1}$$

5.1.2 Standard Consequentialism

A standard consequentialist spectator chooses her allocation criteria and subsequent allocation plan to maximize a value function that depends only on the resulting outcome distribution:

$$V(\pi) = v(\mu^{\pi}). \tag{2}$$

The functional form of $v(\cdot)$ is fully flexible, allowing it to capture a wide range of outcome-based fairness concerns, including preferences for meritocracy, diversity, or compensating low-income

[18] If allocation decisions are made after signals are realized, the spectator may not literally formulate a fully contingent plan in advance. In that case, π should be interpreted as the spectator's possibly incomplete vision of her intended decisions before signals are realized.

candidates. Moreover, it can also accommodate ex-ante notions of fairness such as equality of opportunity. Importantly, the inclusion or exclusion of signals does not directly affect the value function; signals matter only insofar as they shape the induced outcome distribution μ^{π}.

If the chosen allocation plan can be implemented perfectly after signals are realized—either because $V(\pi)$ is dynamically consistent or because the spectator can commit to the plan—then a consequentialist spectator never strictly prefers to exclude any signal from her allocation criteria. The reason is straightforward: if a particular outcome distribution μ is attainable with a smaller set of allocation criteria AC, it is also attainable under any larger AC by simply ignoring the additional signals and choosing the same allocation plan. Hence, the maximum value achievable with a smaller AC is always weakly lower than that attainable with an expanded AC.

5.1.3 Bounded Rationality

If the implementation of the chosen allocation plan is imperfect, then outcome-based fairness concerns can give rise to information exclusion.

One source of imperfect implementation comes from *anticipated decision costs*. The spectator may anticipate mental effort required to formulate and carry out an allocation plan, and this cost c can depend on both the plan π and the set of allocation criteria AC. In this case, the spectator's payoff under AC is

$$\max_{\pi} V(\pi) - c(\pi, AC). \tag{3}$$

Decision costs can induce the spectator to exclude biased or invalid information if including that information increases $c(\pi, AC)$.[19] Such costs may arise if the spectator feels compelled to process all information available, with noise or bias making that effort more demanding. Alternatively, including more signals may increase the likelihood of encountering conflicting signals, and resolving these conflicts in a fair manner can be especially taxing when the spectator must also

[19]For this to happen, the additional cost must arise even when the spectator ultimately disregards the extra signals and implements the same allocation plan she would have chosen without them. If the spectator can costlessly ignore additional information, then the maximum $V(\pi) - c(\pi, AC)$ achievable under a smaller set of criteria is still attainable under any expanded set, and there is no strict preference for information exclusion.

adjust for noise or bias.

Another form of imperfect implementation that may affect information preferences is related to *anticipated decision mistakes*. In this case, the spectator worries that, due to factors such as trembling-hand errors or dynamic inconsistency, the implemented allocation decisions may deviate from the ex-ante plan that maximizes $V(\pi)$. Formally, the payoff under a given set of allocation criteria AC is $V(\tilde{\pi})$, where the actually implemented plan $\tilde{\pi}$ differs from the optimal π.

Anticipated mistakes can lead the spectator to prefer blinding herself to biased or invalid information if she believes such information increases the likelihood or severity of deviations. For instance, if the spectator has a tendency to take realized signals at face value when making decisions, she may insufficiently adjust for noise or bias in the signals. In this case, including biased or invalid information raises the risk that her implemented allocation $\tilde{\pi}$ produces worse outcomes than the decision without the additional information. Anticipating this, the spectator may prefer to exclude the signals altogether.

5.1.4 Procedural Fairness

A spectator may care about the fairness of the allocation *procedure* independent of the resulting outcome distribution. This may be because she follows a deontological rule or moral intuition that requires the procedure to be fair, or because choosing a fair procedure serves as a signal—to herself or to others—of her general commitment to fairness. For spectators with such concerns, the inclusion of invalid or biased information in the allocation criteria may taint the fairness of the entire procedure. Hence, if a spectator cares both about the allocation outcomes and its procedure, the decision to include or exclude a signal depends on the tradeoff between the expected usefulness of that signal (the potential improvement in $V(\pi)$) and the moral cost of including it.

Procedural fairness concerns may also apply to the *use* of information. Even if invalid or biased signals are included in AC, the spectator may prefer an allocation plan π that places less weight on them, since reliance on such signals could compromise the fairness of the procedure.[20]

[20]Formally, procedural fairness concerns about the use of information can be represented by the same functional form as Equation (3), where the cost c increases as the allocation plan π relies more on invalid or biased information

These concerns about use further reduce the value of including invalid or biased signals, as they restrict the efficient utilization of information and thereby limit potential improvements in $V(\pi)$. It is important to note, however, that procedural fairness concerns about information use alone, without a moral cost directly imposed on information inclusion, cannot generate a strict preference for exclusion. This is because even when a signal is included in the allocation criteria, spectators always have the free option to not use it.

5.2 Suggestive Evidence against Outcome-Based Fairness from the Admission Experiment

While the admission experiment is not designed to cleanly separate fairness concerns about admission outcomes and procedures, its findings contain suggestive evidence inconsistent with the former. To begin, recall that spectators' outcome-based social preferences, such as meritocracy and preference for low-income students, are reflected in their admission decisions. However, Tables A3 and A4 show that these preferences, measured by the number of meritocratic decisions and the number of admitted lower-income students for Groups 1 to 3, are uncorrelated with preferences to include or exclude Test 2 scores for any scenario involving test prep.

Moreover, as we discuss in the conceptual framework, for outcome-based fairness concerns to induce information exclusion, they must interact with bounded rationality so that including Test 2 scores would increase either the cost of making admission decisions or the risk of decision mistakes. However, we find several pieces of evidence suggesting that these forms of bounded rationality do not play a key role.

First, very few spectators mention decision costs and decision mistakes as justifications for excluding Test 2 scores. As is shown in Table A8, across all test prep scenarios, only 3% and 4% of exclusion justifications argue that including Test 2 scores would complicate or bias the admission decisions.

Second, if costs of making admission decisions are a key driver of information exclusion, then

in AC, and equals zero if it does not.

spectators who prefer to exclude Test 2 scores should also spend more time on admission decisions when Test 2 scores are available, as decision costs and decision time are often positively correlated (Fudenberg et al., 2018; Halevy et al., 2023). However, as Table A5 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 first 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.[21]

Third, both decision costs and decision mistakes would predict a gap between information preferences when spectators are making admission decisions themselves and when others make these decision. 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. In contrast to both predictions, we find that spectators' advice about whether others should request access to Test 2 scores closely mirrors their own information preferences when making admission decisions themselves (see Appendix B.2 for detailed analysis of advice).

Taken together, the evidence suggests that anticipated decision costs and decision mistakes are not key drivers of the exclusion of Test 2 scores.

5.3 Identifying Procedural Fairness Concerns

While evidence from the admission experiment suggests that concerns about outcome fairness are not the primary driver of information exclusion, we have yet to conclude definitively that exclusion is motivated by procedural fairness concerns. In order to do so, we must be able to vary

[21] An alternative measure of decision costs 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.

the validity and bias of information while holding constant both the distribution of allocation outcomes and the mental costs of making allocation decisions. This is not feasible in the admission experiment, as neither we nor the participants know the precise degrees of noise or bias introduced by test preparation. As a result, we cannot construct equivalent distributions over admission outcomes across different test prep scenarios. To address this limitation, we conduct a more stylized bonus allocation experiment that allows for greater control over key elements of the design.

5.3.1 Bonus Allocation Experiment: Design

Context. We conducted the new experiment on Prolific using a US sample on August 2, 2025 with a participation fee of $6 (see screenshots in Appendix F). Spectators are informed at the outset that a group of workers were recruited to collect online information about US public companies under a flat payment. Each worker spent 5 minutes on each of two tasks, Task 1 focusing on financial firms and Task 2 on service firms. After the tasks ended, an evaluator assigned each worker two task scores, one for each task, to record the number of firms for which accurate information was collected.

Scoring was carried out over two days, Day 1 and Day 2. Each worker was randomly assigned to be scored on one of the two days, so Day 1 and Day 2 workers are ex ante identical in characteristics and performance. The key difference lies in the accuracy of Task 2 scores: while all Day 1 Task 2 scores accurately reflect true performance, a random half of Day 2 workers received an inflated Task 2 score—one point higher than their true score. This random grade inflation renders Task 2 scores invalid for, and biased in favor of, Day 2 workers. By contrast, Task 1 scores are accurate for all workers, regardless of scoring day. Note that the data-generating process of task scores conditional on true performances is fully specified in this design, which is a crucial distinction from the admission experiment. The context is also substantially different, allowing us to examine whether the preference to exclude invalid and biased information generalizes across settings.

Allocation decisions. Spectators' main task is to decide which of two workers should receive an additional \$3 bonus, based on information provided in a report card. Each spectator makes four allocation decisions, structured across two experimental parts, with two conditions per part. The two parts differ in the scoring days of the two candidate workers, which determines the accuracy of their Task 2 scores:

- *Accurate Info* part: Both workers were scored on Day 1, so their Task 2 scores are accurate.
- *Invalid and Biased Info* part: One worker was scored on Day 1 and the other on Day 2. Because only Day 2 scores are subject to random inflation, Task 2 scores in this part are invalid for and biased in favor of the Day 2 worker.

Within each part, spectators make two allocation decisions under different informational conditions:

- *Task 2 Excluded* condition: The report card displays each worker's scoring day and Task 1 score, but omits Task 2 scores.
- *Task 2 Included* condition: The report card displays the scoring day along with both Task 1 and Task 2 scores.

The two experimental parts as well as the two conditions within each part are presented to spectators in random order.

To make an allocation decision for a part-condition, spectators do not select a worker given a specific report card. Instead, they choose and commit to a contingent plan that specifies who receives the bonus under every possible realization of the report card. Table 2 lists the candidate plans available to spectators.[22]

Two features of the choice sets are worth highlighting. First, Plan B—which allocates the bonus solely based on Task 1 scores—is always available, even when Task 2 scores are included on the report card. This feature increases the likelihood that some spectators will choose the same allocation plan across both informational conditions, allowing us to later examine their preferences over the inclusion of Task 2 scores while holding the plan choice fixed.

[22]These plans are presented to spectators in random order.

Condition	*Accurate Info* part	*Invalid and Biased Info* part
Task 2 Excluded	Plan A: Flip a coin to decide who gets the bonus, regardless of their scores. Plan B: Give the bonus to the worker with the higher Task 1 score. If their Task 1 scores are the same, flip a coin to decide.	Plan A Plan B
Task 2 Included	Plan B Plan C: Give the bonus to the worker with the higher total score (Task 1 + Task 2). If their total scores are the same, flip a coin to decide.	Plan B Plan D: Give the bonus to the worker with the higher total score (Task 1 + Task 2). If their total scores are the same, give the bonus to the Day 1 worker.

Table 2: Allocation plans available in each experimental condition and part.

Second, Plan D is constructed to yield the same outcome distribution—that is, the joint distribution of true performance and received bonus—as Plan C, which is arguably the fairest plan, even though it relies on invalid and biased Task 2 scores. To understand this equivalence, note that under Plan D, the Day 2 worker receives the bonus whenever their *true* total performance (Task 1 + Task 2) exceeds that of the Day 1 worker. Conversely, if the Day 1 worker's true total performance is higher, they receive the bonus, since even an inflated Task 2 score for the Day 2 worker can at best create a tie, which is broken in favor of the Day 1 worker. Thus, grade inflation affects Plan D only when the two workers have equal true total performance—in which case, the Day 2 worker receives the bonus if and only if they are randomly assigned the extra point.[23] Since this happens with 50% probability, the Day 2 worker receives the bonus half the time when true performance is tied. Consequently, the bonus distribution under Plan D exactly matches that under Plan C. This equivalence allows us to compare spectators' willingness to *use* invalid and biased information versus accurate information, while holding fixed the consequences of information use. Moreover, because Plan D uses biased information to produce what is arguably the fairest outcome distribu-

[23]This reasoning is explained to the spectators when they choose their plan in the *Invalid and Biased Info* part under the *Task 2 Included* condition.

tion, it enables us to later examine whether spectators still prefer to exclude such information even when the way it is used fully neutralizes its bias.

Implementation of allocation plans. Each spectator is informed that one of their four allocation decisions—corresponding to a specific part-condition—may be randomly selected for real implementation. If selected, we first match the spectator with a pair of workers whose scoring days are consistent with the designated part. We then generate a report card for the two workers, containing the information specified by the selected condition. Finally, the spectator's chosen allocation plan for that part-condition is automatically applied to the report card to determine which worker receives the bonus.

Information preferences. In each part, we elicit spectators' preferences over the inclusion of Task 2 scores in the report card *after* they have chosen and committed to an allocation plan under both informational conditions, with their chosen plans displayed on the screen. This timing ensures that, by the time spectators report their preferences, the outcome distribution associated with each condition is already fixed, and the mental costs of making allocation decisions are sunk. Specifically, we ask whether they would prefer the report card to include or exclude Task 2 scores, or whether they are indifferent. Spectators who express a preference for inclusion or exclusion are given the opportunity to confirm their choice by typing a sentence. Confirming a preference for inclusion (or exclusion) increases the likelihood that Task 2 scores will actually be included (or excluded) on the report card if that part-condition is selected for implementation. Finally, all spectators are asked to provide an open-ended justification for their information preferences.

Logistics. We recruited 200 spectators from Prolific, all of whom resided in the US, had completed at least 200 Prolific surveys, and maintained an approval rate of at least 99%. The median time spent on the survey was 16 minutes. Each spectator received a $6 participation fee.

5.3.2 Bonus Allocation Experiment: Results

Figure 7 shows the distribution of information preferences. Fifty-five percent of spectators prefer to exclude Task 2 scores from the report card when the scores are invalid and biased, and

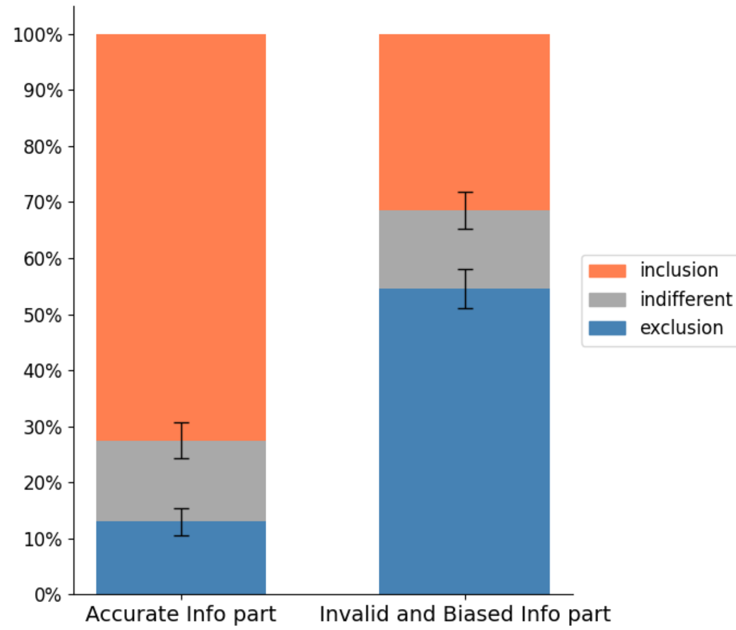


Figure 7: Information preferences in bonus allocation experiment

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

96% of them confirm the strictness of their preference by typing a sentence. In contrast, when the scores are accurate, only 13% of spectators choose to exclude them, with a confirmation rate of 88%. These results are remarkably similar to those from the admission experiment, despite substantial differences in experimental design.

Moreover, as Table 3 shows, 82% $(86/105)$ of spectators who strictly prefer to exclude invalid and biased Task 2 scores had already committed to Allocation Plan B—that is, basing their bonus allocation decisions solely on Task 1 scores—regardless of whether Task 2 scores appear on the report card. As a result, their exclusion preferences cannot be attributed to differences in allocation outcomes. Nor can they be explained by the mental costs of making allocation decisions, since these decisions are made *before* spectators report information preferences, making those costs sunk. Taken together, these findings indicate that the overwhelming majority of preferences to exclude invalid and biased information are driven by non-consequentialist fairness concerns about the allocation procedure itself.

Plan choice		*Accurate Info* part		*Invalid and Biased Info* part	
Task 2 Excluded	*Task 2 Included*	Total	Strictly prefer exclusion	Total	Strictly prefer exclusion
A	B	2	0	18	9
B	B	19	6	116	86
A	C/D	14	3	6	0
B	C/D	165	14	60	10
Total		200	23	200	105

Table 3: Allocation plan choices and preferences for excluding Task 2 scores

Notes: This table reports the distribution of allocation plan choices across experimental parts. For each part, "Total" indicates the number of spectators who selected a given combination of plans under the *Task 2 Excluded* and *Task 2 Included* conditions. "Strictly prefer exclusion" shows how many of these spectators expressed a strict preference for excluding Task 2 scores from the report card.

Spectators not only prefer to exclude invalid and biased scores from the allocation criteria, but also exhibit aversion to *using* them even when they are available on the report card. Only 33% $(66/200)$ of spectators choose Plan D—which allocates the bonus based on total scores (Task 1 + Task 2)—over Plan B, which relies solely on Task 1 scores. This is in stark contrast ($p < 0.001$) to the 90% $(179/200)$ of spectators who choose to use total scores under Plan C, despite the fact that Plans C and D yield the same, and arguably fairest, distribution of allocation outcomes.[24] Moreover, preferences for using and including invalid and biased Task 2 scores are highly positively correlated ($\rho = 0.623$, $p < 0.001$). These results suggest that the *use* of invalid and biased information is influenced by the same type of non-consequentialist fairness concerns that underlie preferences over its inclusion in the allocation criteria.

[24]This result mirrors the finding from the admission experiment reported in Section 3.2.2, where spectators are less likely to admit students based on Test 2 scores when the scores are biased, even when the bias does not affect the students under consideration.

6 Conclusion and Discussion

This paper experimentally documents a prevalent and robust preference for excluding invalid or biased information from allocation criteria. In our main experiment on admission decisions, we show that the inclusion or exclusion of invalid or biased test scores reflects a tradeoff between their instrumental usefulness and fairness concerns about incorporating them. In our auxiliary experiment on bonus allocations, we further show that these fairness concerns are primarily non-consequentialist and procedural in nature: participants prefer to exclude invalid or biased information even when allocation outcomes are held constant across information conditions.

6.1 Psychology behind Non-consequentialist Preference for Information Exclusion

The preference to exclude invalid or biased information for non-consequential reasons could arise from moral intuitions or deontological rules, or from the desire to signal concern for fairness to oneself or others. These explanations are often difficult to disentangle (see, e.g., Henkel et al., 2024), and doing so lies beyond the scope of this paper. Nonetheless, several of our findings suggest that the procedural fairness concerns documented here are unlikely to be driven primarily by conscious signaling.

First, if exclusion served mainly as a signaling strategy, it should be more common when information preferences are observed by more people. Yet we find no significant difference between participants' own information preferences—which are only visible to the experimenters—and the advice they provide to other spectators, which has some chance of being observed more widely. Second, if exclusion were primarily a way to signal disapproval of invalidity or bias, it should be less common among spectators in the bonus allocation experiment who had already committed to an allocation plan that did not use biased information, since this commitment itself constitutes a strong signal of disapproval. However, we find that the vast majority of these spectators still strictly prefer to exclude invalid and biased scores. Taken together, these results suggest that procedural concerns about including invalid or biased information are more plausibly rooted in moral intuition or deontological commitments than in strategic signaling.

6.2 External Relevance for Test Policies in College Admissions

While our primary goal is to provide a proof of concept and investigate the mechanisms behind the idea that fairness concerns can lead people to prefer excluding biased or invalid information from allocation criteria, results from our main admission experiment also shed light on why test-optional and test-blind college admission policies have gained substantial public support.

First, within our experiments, the preference to exclude biased or invalid test scores is robust to changes in the broader information environment and to who makes the admission decisions. This robustness suggests that the exclusion preference is not an artifact of any specific design choice we make, but instead reflects a stable judgment principle that is likely to generalize beyond the lab.

Second, evidence from both outside and inside the lab points in the same direction. In Appendix C, we analyze a national public opinion poll and find that support for test-optional and test-blind policies is strongly correlated with perceptions that standardized tests are invalid and biased. Within our own admission experiment, when participants are asked about real-world test policies, they highlight the fairness and usefulness of standardized tests as their primary considerations (Appendix E). Moreover, participants who invoke fairness when justifying their experimental information preferences are significantly more likely to invoke fairness when reasoning about real-world test policies. This convergence between survey data and experimental evidence suggests that the same fairness-driven exclusion mechanism we identify in the lab also helps explain public support for test-optional and test-blind policies.

Of course, our experiments abstract away from many important features of real-world admissions that may shape attitudes toward test policies. Beyond differences in test preparation, spectators pointed to test-taking skills and comparability across applicants as factors influencing their attitudes toward standardized test scores. Moreover, the most common admissions policy today is not test-required or test-blind but test-optional, which allows applicants to choose whether to report their scores. Although spectators' attitudes toward test-optional and test-blind policies are highly correlated, they also highlight considerations unique to test-optional admissions, such as the flexibility and autonomy afforded to applicants, which increase their support. At the same time, recent studies show that unequal ability to take advantage of this discretion can create unintended disparities in outcomes (see, e.g., Exley et al., 2024). A systematic investigation of how these features affect attitudes toward standardized tests and test policies in admissions is an important direction for future research.

6.3 Broader Implications

More broadly, our results have implications for the design of fair allocation mechanisms. The computer science literature on algorithmic fairness largely emphasizes equalizing error rates of algorithmic predictions—inputs to allocation decisions—across demographic groups (see, e.g., Berk et al., 2021). Achieving such parity often requires excluding predictive but biased variables from training data (see, e.g., Yang and Dobbie, 2020), a practice that Rambachan et al. (2020) show is inconsistent with consequentialist principles. Our findings suggest that procedural fairness concerns may help explain why exclusion remains a popular approach in algorithmic fairness debates. They also highlight an important constraint for mechanism design: allocation criteria must not only produce good outcomes but also be perceived as procedurally fair.

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.

I. Almås, A. W. Cappelen, and B. Tungodden. Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy*, 128(5):1753–1788, 2020.

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.

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.

R. Berk, H. Heidari, S. Jabbari, M. Kearns, and A. Roth. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1):3–44, 2021.

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

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.

C. Buchmann, D. J. Condron, and V. J. Roscigno. Shadow education, american style: Test preparation, the sat and college enrollment. *Social forces*, 89(2):435–461, 2010.

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.

J. Dana, R. A. Weber, and J. X. Kuang. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33:67–80, 2007.

W. Dessein, A. Frankel, and N. Kartik. The test-optional puzzle. *AEA Papers and Proceedings*, 115:669–675, 2025.

W. Dessein, A. Frankel, and N. Kartik. Test-optional admissions. *American Economic Review*, forthcoming.

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 J. B. Kessler. Information avoidance and image concerns. *The Economic Journal*, 133(656):3153–3168, 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`.

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. Beyond instrumental value: How complexity shapes information demand. 2025.

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.

L. Henkel, R. Bénabou, A. Falk, and J. Tirole. Eliciting moral preferences under image concerns: Theory and experiment. Technical report, CRC TR 224 Discussion Paper Series, 2024.

M. Hoffman. The value of hiring through employee referrals in developed countries. *IZA World of Labor*, 2017.

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

M. Minow, J. Zittrain, J. Bower, et al. Technical flaws of pretrial risk assessments raise grave concerns, 2019.

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

E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, et al. A large-scale analysis of racial disparities in police stops across the united states. *Nature human behaviour*, 4(7):736–745, 2020.

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

S. Saccardo and M. Serra-Garcia. Enabling or limiting cognitive flexibility? evidence of demand for moral commitment. *American Economic Review*, 113(2):396–429, 2023.

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

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

Online Appendix

A Additional Figures and Tables

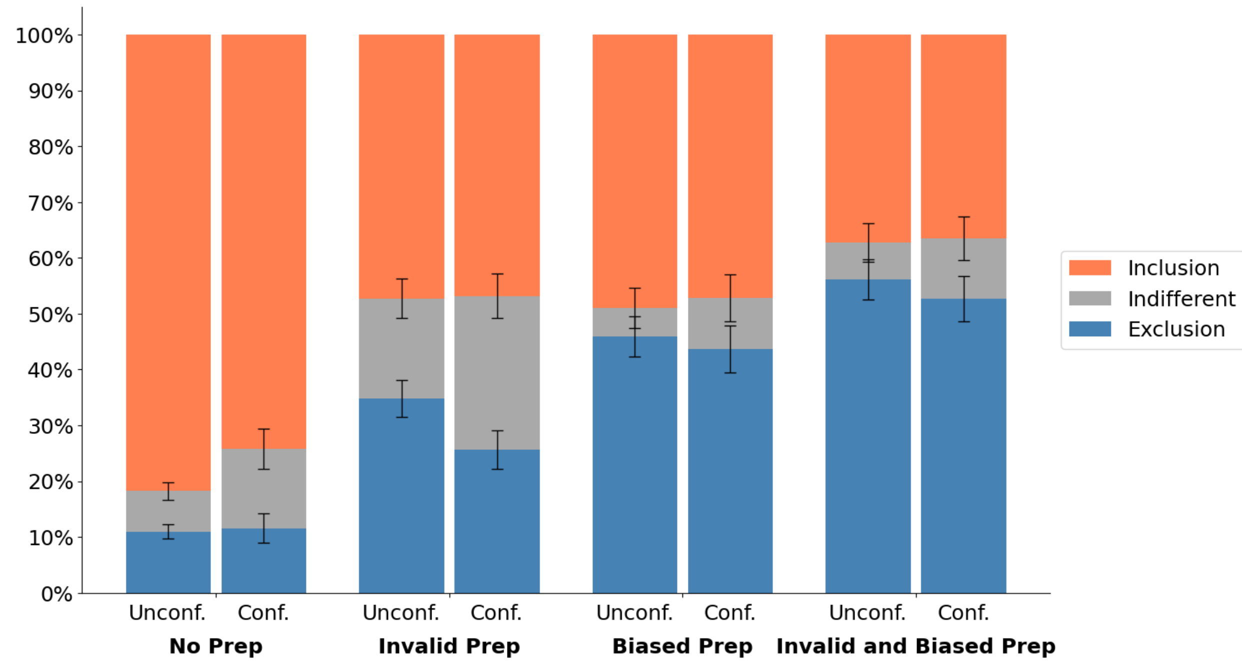


Figure A1: Confirmed information preferences across test prep scenarios (before admission experience)

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.

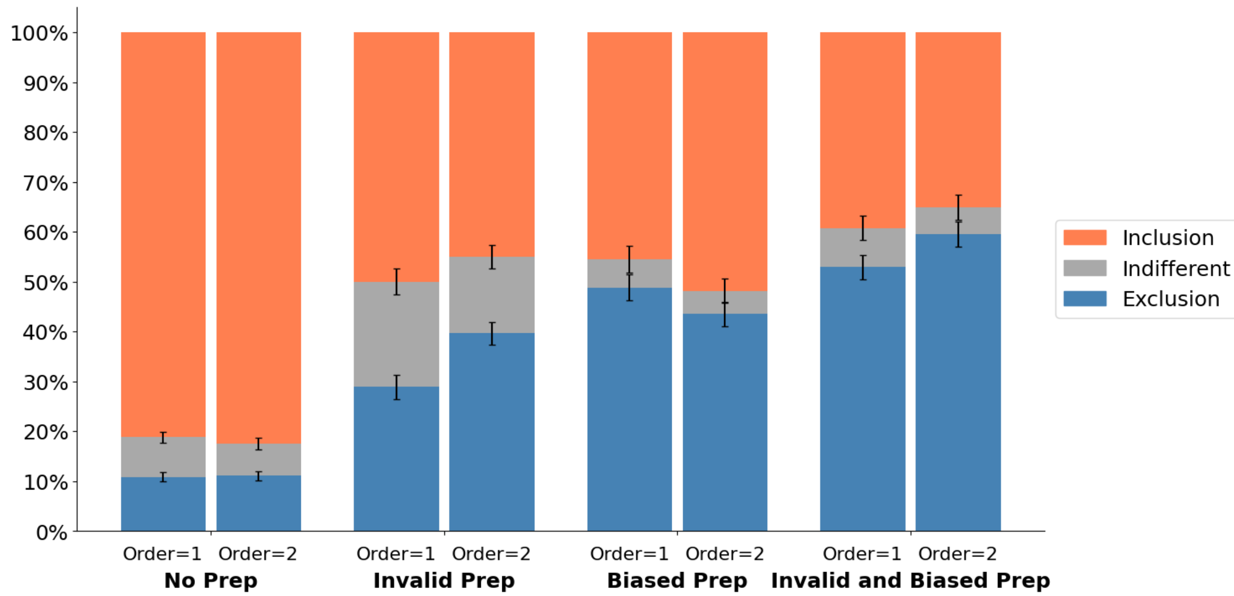


Figure A2: Information preferences by the order between test prep scenarios (before admission experience)

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.

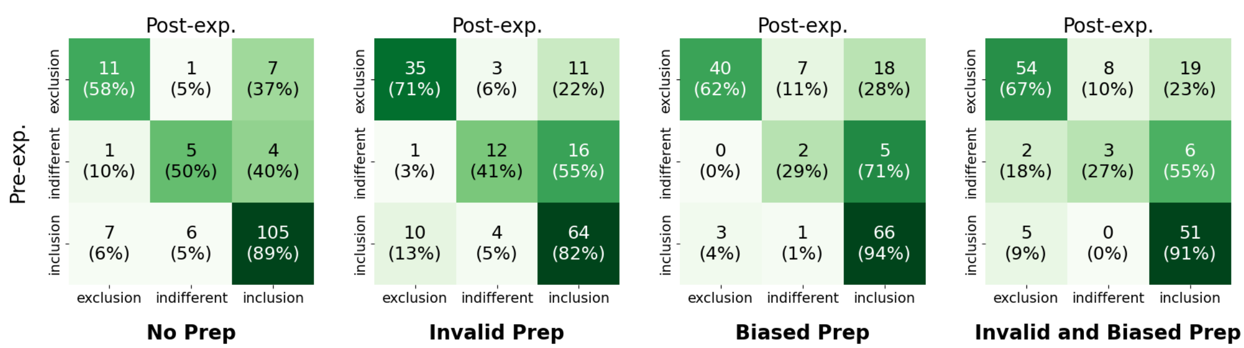


Figure A3: Transition matrix for information preferences before and after admission experience

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.

	Information Preferences			
	OLS&FE	OLS	oLogit&FE	oLogit
Test prep scenario (ref: No Prep)				
Invalid Prep	-0.520***	-0.601***	-1.575***	-1.558***
	(0.075)	(0.071)	(0.285)	(0.167)
Biased Prep	-0.687***	-0.633***	-1.567***	-1.599***
	(0.089)	(0.081)	(0.251)	(0.200)
Invalid & Biased Prep	-0.950***	-0.908***	-2.384***	-2.199***
	(0.081)	(0.076)	(0.324)	(0.192)
Test prep scenario order	-0.022	-0.027	-0.128	-0.060
	(0.048)	(0.050)	(0.163)	(0.137)
Age bracket		0.054**		0.148*
		(0.023)		(0.063)
Gender = Male		0.008		0.027
		(0.054)		(0.147)
Race = White		0.058		0.149
		(0.060)		(0.165)
Employment status (ref: Working full-time)				
Working part-time		0.070		0.230
		(0.072)		(0.199)
Unemployed and looking for work		0.104		0.291
		(0.099)		(0.276)
Homemaker or stay-at-home parent		0.038		0.113
		(0.101)		(0.280)
Student		0.100		0.301
		(0.143)		(0.354)
Retired		-0.038		-0.092
		(0.117)		(0.326)
Education Level		0.030		0.085
		(0.020)		(0.056)
Liberal orientation		0.016		0.040
		(0.022)		(0.061)
Income bracket		0.023		0.069
		(0.018)		(0.049)
Spectator FE	Yes	No	Yes	No
R^2	0.58	0.18		
Pseudo R^2			0.43	0.11
Observations	1,186	1,084	664	1,084

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

Notes: This table shows OLS and ordered logit 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. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

	Post-experience Information Preference			
	No Prep	Invalid Prep	Biased Prep	Invalid and Biased Prep
# of Test-2-focused decisions	0.61**	0.63**	0.58**	0.60**
	(0.23)	(0.21)	(0.21)	(0.20)
Pre-exp. information preference (ref: exclude)				
indifferent	0.58	2.24***	1.97*	1.42*
	(0.76)	(0.49)	(0.89)	(0.64)
include	2.68***	2.93***	3.61***	3.42***
	(0.58)	(0.46)	(0.59)	(0.54)
Category thresholds oLogit				
1	0.86	1.61	1.13	1.38
	(0.57)	(0.43)	(0.37)	(0.34)
2	1.66	2.46	1.72	1.89
	(0.59)	(0.46)	(0.40)	(0.36)
Pseudo R^2	0.21	0.24	0.35	0.29
Observations	147	156	142	148

Table A2: Ordered Logit analysis of the use of Test 2 scores and post-experience information preferences

Notes: This table shows the ordered logit estimates of regressing information preferences after admission experience on the number of Test-2-focused decisions for Groups 4 to 6, controlling for information preferences before admission experience. Post-experience information preferences are coded as follows: include: 1; indifferent: 0; exclude: -1. Pre-experience information preferences are coded as categorical variables (reference = exclude). 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. $^{***}: p < 0.001$; $^{**}: p < 0.01$; $^{*}: p < 0.05$

	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 A3: 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. $^{***}: p < 0.001$; $^{**}: p < 0.01$; $^{*}: p < 0.05$

	Information Preference (include: 1; indifferent: 0; exclude: -1)			
	No Prep	Invalid Prep	Biased Prep	Invalid and Biased Prep
# of meritocratic decisions	0.27	-0.01	-0.16	-0.39
	(0.15)	(0.22)	(0.20)	(0.21)
# of lower-income admissions	0.09	0.11	-0.07	-0.12
	(0.08)	(0.10)	(0.10)	(0.11)
Category thresholds oLogit				
1	-1.16	-0.29	-0.77	-1.06
	(0.54)	(0.80)	(0.71)	(0.79)
2	-0.56	0.45	-0.56	-0.78
	(0.54)	(0.80)	(0.71)	(0.79)
Pseudo R^2	0.01	0.01	0.002	0.01
Observations	593	201	196	196

Table A4: Ordered logit analysis of information preferences (before admission experience) and outcome-based social preferences

Notes: This table shows ordered logit 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. $^{***}: p < 0.001$; $^{**}: p < 0.01$; $^{*}: p < 0.05$

	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	148

Table A5: 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. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

B Additional Analysis

B.1 The Preference to Exclude Invalid and Biased Test Scores is Robust to the Availability and Quality of 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. [25] 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.[26]

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

[25] 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 other hand, FairTest, a non-profit organization, justified test-optional and test-blind policies by citing a study that the grades are quite informative predictors especially within demographic categories (Feder and Bello, 2024).

[26] 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).

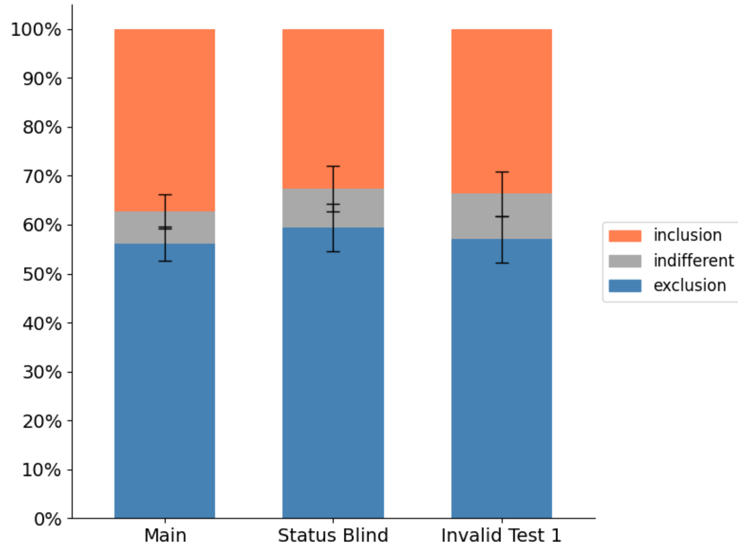


Figure A4: Information preferences across experiments: Invalid and Biased Prep scenario

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.

cannot use other applicant information to adjust for the scores' bias.

Figure A4 shows that the demand for Test 2 scores in these two treatments is virtually the same as the main treatment. This result indicates that although the value of test scores should depend on other available student information, people tend to evaluate the scores in isolation when they consider their demand.

B.2 Advice for Others is Consistent with Own Information Preferences

In the admission experiment, after six rounds of admission decisions, we ask spectators to advise other participants on whether they should request access to Test 2 scores when facing the same test prep scenario. Figure A5 cross-tabulates the advice with spectators' own information preferences elicited right before the advice for each test prep scenario. 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, nearly half (6) of these reversals come from the No Prep scenario, suggesting that decision costs are not

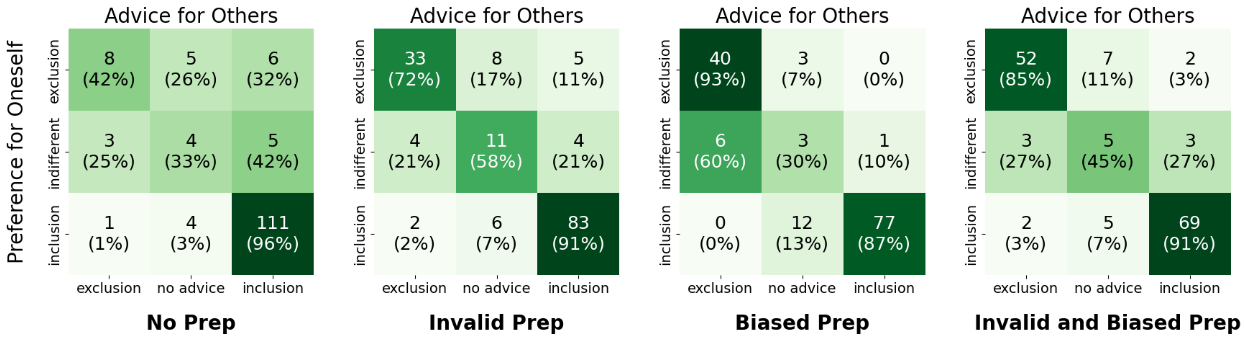


Figure A5: Transition matrix for information preferences for oneself and advice for others across test prep scenarios

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.

a main reason for excluding invalid or biased Test 2 scores. 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 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 (40%) of this behavior is justified by respect for others' agency (see Appendix D.3 for a full classification of advice justifications). Among those who express indifference for themselves, 25% advise for inclusion and 31% advise against it.

The strong consistency between spectators' own information preferences and the advice they give to others suggests that these preferences do not depend heavily on who makes the admission decisions. This finding enhances the external validity of our study, as many people who hold opinions about college admissions policies are not actual decision makers themselves. Moreover, the absence of a gap between own preferences and advice provides suggestive evidence against explanations of information exclusion based on anticipated decision costs or decision mistakes—both of which would predict a gap between the two measures (see Section 5.2 for further discussion).

C Attitudes toward Standardized Test Requirement and Perceptions about Standardized Test in a Harris Poll Survey

To corroborate our finding that perceptions of validity and bias of test scores drive attitudes toward test policies in college admissions, we analyze data from a public opinion poll on standardized tests conducted by the Harris Poll in 2021.[27] Table A6 shows the result of regressing attitudes toward test requirement in college admissions on various perceptions about standardized tests. The result shows that support for test-optional and test-blind admission policies is highly correlated with perceptions that the scores are invalid and biased.

	Agree to stop requiring standardized test scores
High school grades are better measure than standardized test scores	0.29***
	(0.03)
Standardized tests help low-income students with low GPA	0.06**
	(0.03)
Standardized tests biased against Black and Hispanic students	0.17***
	(0.04)
Standardized tests biased in favor of White and Asian students	0.06
	(0.04)
Standardized tests biased in favor of affluent students	0.18***
	(0.03)
Standardized tests as objective measure for comparison across US	-0.04
	(0.03)
Standardized tests predict college performance	-0.04
	(0.03)
Standardized tests measure academic knowledge/skills	-0.12***
	(0.04)
R^2	0.32
Observations	1,000

Table A6: Regression analysis of attitudes toward standardized test requirement and perceptions about standardized tests in a Harris Poll survey

Notes: This table shows OLS estimates of regressing attitudes toward standardized test requirement on perceptions of standardized tests using raw data from a public opinion poll conducted by the Harris Poll in 2021. All variables are measured on a 1–4 scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, and 4 = strongly agree. Numbers in parentheses are standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

[27] https://theharrispoll.com/briefs/americans-want-to-end-standardized-tests-for-good/

D Self-reported Rationales for Test 2 Scores Exclusion and Inclusion

We use GPT-3.5, a large language model, to summarize spectators' open-ended justifications for their information preferences. The summarization procedure and its results are detailed in Appendix D.1. To validate these findings, we also hand-code each justification to identify the considerations mentioned, using the coding scheme described in Appendix D.2. The same scheme is applied to spectators' justifications for their advice on whether others should request access to Test 2 scores. The results of the coding exercise are reported in Appendix D.3.

D.1 Procedure for Summarizing Open-ended Rationales

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, compromis-

ing 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 summary 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.

D.2 Coding Scheme

Code	Explanation	Example
Mention Test 2 score usefulness	Mention Test 2 scores' usefulness in making admissions decisions.	"The score will still be able to help me determine which students perform the best even with the test prep."
Mention Test 2 score fairness	Mention that Test 2 score could be an unfair measure, and/or mention that including Test 2 scores makes the evaluation unfair.	"Since test prep was received, it would not be a fair comparison between the two groups to use their test 2 scores as part of the admission process." "It's unfair that one group should have an advantage that the other does not enjoy."
Test 2 scores bias admissions	Including Test 2 scores could introduce bias and mistakes into admissions decisions.	"[...]Including those scores would bias the reviewers in favor of higher-income students."
Test prep makes scores invalid	Test prep can distort the validity of Test 2 scores as measures of ability.	"As the test prep didn't improve students' data analysis skills, relying solely on Test 2 scores could paint an inaccurate picture of their overall understanding of the subject matter."

Code	Explanation	Example
Can adjust for test prep	Test prep effects can be quantified and adjusted for in scoring.	“I feel like I can adjust for the fact that they may have received test prep.”
Can extract information amid bias	Meaningful score comparisons are still possible within income groups and for large score differences across groups.	“It only may have boosted their Test 2 scores by up to 2 points. any difference larger than 2 is still informative.”
Test 2 scores break-tie for Test 1	Test 2 scores can differentiate between students with identical Test 1 scores.	“If there is a tie for spots in the class from test 1, test two will help break it.”
Want to compare Test 1 & 2	Desire to evaluate consistency between Test 1 and Test 2 performance.	“Even with the help the higher income received in test 2, I want to see their scores in order to see if they are consistent with test 1 where they didn’t receive any help.”
Want more information	Desire to include Test 2 scores as additional data.	“I want as much information as is available.”

Code	Explanation	Example
Curiosity	Curious to see Test 2 scores.	"I'm curious to see whether higher income students with prep fared better than those without."
Include Test 2 for transparency	Including Test 2 scores to maintain transparency about student performance.	"Including Test 2 scores on the report cards offers transparency on the students' performance and any potential impact from the preparation."

Table A7: Classification of Open-Ended Rationales on Whether to Include or Exclude Test 2 Scores in the Experiment

Notes: This table presents an overview of the coding scheme for self-reported rationales on whether to include or exclude Test 2 scores in the experiment. It includes category labels, explanations for each code, and example excerpts from open-text responses that belong to each category.

D.3 Full Classifications

Reasons for exclusion	No Prep (N=65)	Invalid Prep (N=70)	Biased Prep (N=90)	Invalid & Biased Prep (N=110)
Mention Test 2 score usefulness	11 (17%)	58 (83%)	24 (27%)	39 (35%)
Mention Test 2 score fairness	3 (5%)	7 (10%)	73 (81%)	74 (67%)
Test 2 scores bias admissions	0 (0%)	5 (7%)	1 (1%)	7 (6%)
Test 2 scores complicate admissions	4 (6%)	1 (1%)	1 (1%)	4 (4%)
Test prep makes scores invalid	NA	51 (73%)	16 (18%)	31 (28%)
Can adjust for test prep	NA	0 (0%)	0 (0%)	0 (0%)
Can extract information amid bias	NA	1 (1%)	0 (0%)	0 (0%)
Test 2 scores break-tie for Test 1	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Want to compare Tests 1 & 2	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Want more information	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Curiosity	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Include Test 2 for transparency	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table A8: Reasons for Excluding Test 2 Scores by Test Prep Scenario (Pre-experience)

Notes: This table shows the frequencies of spectators' reasons for preferring to exclude Test 2 scores under each test prep scenario. We include categories suggested by the theories in Section 5.1 and topics summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

Reasons for inclusion	**No Prep**	**Invalid Prep**	**Biased Prep**	**Invalid & Biased Prep**
	(N=485)	**(N=95)**	**(N=96)**	**(N=73)**
Mention Test 2 score usefulness	290 (60%)	55 (58%)	47 (49%)	40 (55%)
Mention Test 2 score fairness	104 (21%)	8 (8%)	12 (13%)	11 (15%)
Test 2 scores bias admissions	0 (0%)	0 (0%)	1 (1%)	0 (0%)
Test 2 scores complicate admissions	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Test prep makes scores invalid	NA	1 (1%)	1 (1%)	1 (1%)
Can adjust for test prep	NA	4 (4%)	11 (11%)	8 (11%)
Can extract information amid bias	NA	16 (17%)	17 (18%)	10 (14%)
Test 2 scores break-tie for Test 1	4 (1%)	0 (0%)	0 (0%)	0 (0%)
Want to compare Tests 1 & 2	38 (8%)	5 (5%)	3 (3%)	5 (7%)
Want more information	140 (29%)	30 (32%)	17 (18%)	23 (32%)
Curiosity	9 (2%)	6 (6%)	13 (14%)	4 (5%)
Include Test 2 for transparency	6 (1%)	5 (5%)	6 (6%)	7 (10%)

Table A9: Reasons for Including Test 2 Scores by Test Prep Scenario (Pre-experience)

Notes: This table shows the frequencies of spectators' reasons for preferring to include Test 2 scores under each test prep scenario. We include categories suggested by the theories in Section 5.1 and topics summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

Reasons for indifference	No Prep (N=43)	Invalid Prep (N=36)	Biased Prep (N=10)	Invalid & Biased Prep (N=13)
Mention Test 2 score usefulness	9 (21%)	16 (43%)	3 (30%)	2 (15%)
Mention Test 2 score fairness	0 (0%)	7 (19%)	1 (10%)	2 (15%)
Test 2 scores bias admissions	0 (0%)	0 (0%)	0 (0%)	1 (8%)
Test 2 scores complicate admissions	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Test prep makes scores invalid	NA	12 (33%)	1 (10%)	2 (15%)
Can adjust for test prep	NA	0 (0%)	0 (0%)	0 (0%)
Can extract information amid bias	NA	0 (0%)	2 (20%)	0 (0%)
Test 2 scores break-tie for Test 1	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Want to compare Tests 1 & 2	0 (0%)	1 (3%)	0 (0%)	0 (0%)
Want more information	1 (2%)	2 (6%)	1 (10%)	1 (8%)
Curiosity	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Include Test 2 for transparency	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table A10: Reasons for Indifference about the Inclusion of Test 2 Scores by Test Prep Scenario (Pre-experience)

Notes: This table shows the frequencies of spectators' reasons for being indifferent about the inclusion of Test 2 scores under each test prep scenario. We include categories suggested by the theories in Section 5.1 and topics summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

Reasons for exclusion	**No Prep (N=12)**	**Invalid Prep (N=39)**	**Biased Prep (N=46)**	**Invalid & Biased Prep (N=57)**
Mention Test 2 usefulness	6 (50%)	27 (69%)	6 (13%)	19 (33%)
Mention Test 2 fairness	2 (17%)	0 (0%)	34 (74%)	29 (51%)
Test 2 scores bias admissions	0 (0%)	1 (3%)	4 (9%)	3 (5%)
Test 2 scores complicate admissions	3 (25%)	1 (3%)	2 (4%)	2 (4%)
Test prep makes scores invalid	NA	25 (64%)	6 (13%)	18 (32%)
Can adjust for test prep	NA	0 (0%)	2 (4%)	3 (5%)
Can extract information amid bias	NA	0 (0%)	0 (0%)	0 (0%)
Test 2 scores break-tie for Test 1	0 (0%)	0 (0%)	1 (2%)	0 (0%)
Want to compare Tests 1 & 2	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Want more information	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Curiosity	0 (0%)	1 (3%)	0 (0%)	0 (0%)
Include Test 2 for transparency	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other spectators have agency	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table A11: Reasons for Recommending to Exclude Test 2 Scores by Test Prep Scenario (Advice for others)

Notes: This table shows the frequencies of spectators' reasons for advising others to exclude Test 2 scores under each test prep scenario. We include categories suggested by the theories in Section 5.1 and topics summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

Reasons for inclusion	**No Prep (N=122)**	**Invalid Prep (N=92)**	**Biased Prep (N=78)**	**Invalid & Biased Prep (N=74)**
Mention Test 2 usefulness	92 (75%)	68 (74%)	61 (78%)	49 (66%)
Mention Test 2 fairness	9 (7%)	2 (2%)	1 (1%)	3 (4%)
Test 2 scores bias admissions	0 (0%)	2 (2%)	0 (0%)	0 (0%)
Test 2 scores complicate admissions	1 (1%)	1 (1%)	0 (0%)	0 (0%)
Test prep makes scores invalid	NA	1 (1%)	0 (0%)	0 (0%)
Can adjust for test prep	NA	10 (11%)	11 (14%)	16 (22%)
Can extract information amid bias	NA	5 (5%)	19 (24%)	8 (11%)
Test 2 scores break-tie for Test 1	6 (5%)	7 (8%)	3 (4%)	2 (3%)
Want to compare Tests 1 & 2	20 (16%)	7 (8%)	14 (18%)	8 (11%)
Want more information	35 (29%)	31 (34%)	21 (27%)	19 (26%)
Curiosity	1 (1%)	2 (2%)	4 (5%)	5 (7%)
Include Test 2 for transparency	0 (0%)	2 (2%)	1 (1%)	1 (1%)
Other spectators have agency	0 (0%)	1 (1%)	0 (0%)	1 (1%)

Table A12: Reasons for Recommending to Include Test 2 Scores by Test Prep Scenario (Advice for others)

Notes: This table shows the frequencies of spectators' reasons for for advising others to include Test 2 scores, under each test prep scenario. We include categories suggested by the theories in Section 5.1 and topics summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

Reasons for giving no advice	**No Prep (N=13)**	**Invalid Prep (N=25)**	**Biased Prep (N=18)**	**Invalid & Biased Prep (N=17)**
Mention Test 2 usefulness	0 (0%)	1 (4%)	2 (11%)	5 (29%)
Mention Test 2 fairness	0 (0%)	0 (0%)	2 (11%)	1 (6%)
Test 2 scores bias admissions	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Test 2 scores complicate admissions	1 (8%)	0 (0%)	0 (0%)	0 (0%)
Test prep makes scores invalid	NA	1 (4%)	0 (0%)	1 (6%)
Can adjust for test prep	NA	0 (0%)	0 (0%)	1 (6%)
Can extract information amid bias	NA	0 (0%)	0 (0%)	1 (6%)
Test 2 scores break-tie for Test 1	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Want to compare Tests 1 & 2	0 (0%)	1 (4%)	0 (0%)	0 (0%)
Want more information	1 (8%)	0 (0%)	1 (6%)	0 (0%)
Curiosity	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Include Test 2 for transparency	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other spectators have agency	3 (23%)	7 (28%)	7 (39%)	7 (41%)

Table A13: Reasons for Giving No Advice on Test 2 Scores Inclusion by Test Prep Scenario (Advice for others)

Notes: This table shows the frequencies of spectators' reasons for for providing no advice to others regarding the inclusion of Test 2 scores, under each test prep scenario. We include categories suggested by the theories in Section 5.1 and topics summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

E Attitudes toward Test Policies in the Real World

E.1 Summary of Results

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 E.2 for details of the summarizing procedure). We then manually classify each spectator's rationale in terms of whether it mentions these two concerns. 24% and 64% of rationales mention the fairness and usefulness of standardized test scores, respectively.[28] 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 ($r = 0.10$, $p = 0.01$).

Nevertheless, while supports for test-blind and test-optional admission policies in our survey are highly correlated with each other ($r = 0.70$, $p < 0.001$), they do not significantly correlate with preferences for excluding Test 2 scores in any test prep scenario in the experiment. 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 6% of spectators), spectators also mention differential test-taking skills (14%) and the across-

[28] Appendix D.3 shows the full classification of rationales.

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.

E.2 Procedure for Summarizing Open-ended Rationales

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.

E.3 Coding Scheme

Code	Explanation	Example
Mention usefulness of SAT/ACT scores	Mention SAT/ACT scores' usefulness in making admissions decisions.	"[...] Test scores can provide a good snapshot on that student's intelligence and general knowledge."
Mention fairness of SAT/ACT scores	Mention that SAT/ACT scores could be unfair measures for students, and/or mention that including SAT/ACT scores in admissions is unfair.	"Academic testing such as SAT or ACT tests can be racially biased. "
Making SAT/ACT optional complicates decisions	Admissions decisions become harder when some applicants submit SAT scores and others do not.	"It seems like mixing the test data would make decision making more complex."
Making SAT/ACT optional biases decisions	Making tests optional creates bias against students who don't submit scores.	"I think making it optional would only lead to more biases since they will only be submitted with high scores."
Test prep affects scores validity	Test prep can distort the validity of SAT/ACT scores as measures of ability.	"High income students often have access to test prep and test prep raises scores on standardized tests. The scores are often raised because students know tricks to getting a higher score."

Code	**Explanation**	**Example**
Mention other determinants of score validity	Other factors (e.g., test stress) can distort the validity of SAT/ACT scores as measures of ability	"[...] There are some who are super smart but aren't good at taking tests so therefore the act and sat would not show how smart they really are."
More information is better	Including SAT/ACT scores as additional data is better.	"It is simply more data to help make an informed decision of who to admit to a limited amount of spaces available. "
SAT/ACT could be down-weighted	SAT/ACT scores can be given less weight in evaluation.	"I feel like tests are a key point but they should just be weighted less."
SAT/ACT could be viewed in context	SAT/ACT scores can be interpreted considering the context.	"I guess admission officers will be able to judge a student's family income when comparing test scores and be able to reduce bias in scores."
Students should be evaluated holistically	Students should be evaluated through a holistic review.	"[...] In reality we should take a holistic view of the students' entire educational experience."
SAT/ACT has a consistent standard	SAT/ACT provides standardized comparison across different student backgrounds.	"I think universities need some universal metric by which to evaluate students. High school grades can be easily manipulated."

Code	Explanation	Example
Other admission criteria are more unfair	Other application components may have greater bias than standardized tests.	"While test-prep is helpful for these tests, advantaging higher-income students, all other aspects of college admissions are advantaged by income as well and are more easily gained/even faked for rich students (e.g. essays, letters of recommendation)."
Tests are stressful	Required testing creates student stress.	"SAT and ACT exams put a lot of pressure on students. "
Test requirement can motivate hard work	Required testing encourages students to work harder.	"It's a goal for kids to work toward."
More freedom to submit/withhold scores	Allowing students to choose whether to submit their scores gives them flexibility and freedom.	"I like this option better because it allows the students to have a choice of if they are going to submit it or not."
Test-optional helps enhance applications	Optional submission allows students to strengthen their applications through selective reporting.	"If you've got a good score you worked hard for, you should be able to include it, but if it's bad I don't think you should be required to. [...] Show your best self in whatever way makes you look best."

Code	Explanation	Example
Same criteria for all applicants	All applicants should be evaluated using identical criteria.	“I think criteria should be the same for all. So it’s every tests or no one tests.”
Making SAT/ACT optional increases applications	Test-optional policy increases application pool.	“[...] A test-optional policy promotes inclusivity by accommodating students from diverse backgrounds who may face challenges in standardized testing.”

Table A14: Classification of Open-Ended Rationales Regarding Attitudes towards Test-Blind and Test-Optional Admissions Policies

Notes: This table presents an overview of the coding scheme for self-reported rationales regarding attitudes toward test-blind and test-optional admissions policies. It includes category labels, explanations for each code, and example excerpts from open-text responses that belong to each category.

E.4 Full Classifications

Reason	Support Test-Blind (N=211)	Indifferent (N=114)	Support Test-Required (N=268)
Mention usefulness of SAT/ACT scores	115 (55%)	18 (16%)	198 (74%)
Mention fairness of SAT/ACT scores	83 (39%)	6 (5%)	39 (15%)
Requiring SAT/ACT complicates decisions	0 (0%)	0 (0%)	2 (1%)
Requiring SAT/ACT biases decisions	1 (0%)	2 (1.75%)	0 (0%)
Test prep makes scores invalid	30 (14%)	1 (1%)	4 (1%)
Mention other determinants of score validity	106 (50%)	18 (16%)	148 (55%)
More information is better	1 (0%)	2 (2%)	13 (5%)
SAT/ACT could be down-weighted	9 (4%)	4 (4%)	14 (5%)
SAT/ACT could be viewed in context	0 (0%)	0 (0%)	3 (1%)
Students should be evaluated holistically	5 (2%)	0 (0%)	0 (0%)
SAT/ACT has a consistent standard	1 (0%)	3 (3%)	21 (8%)
Other admission criteria are more unfair	0 (0%)	1 (1%)	8 (3%)
Tests are stressful	9 (4%)	0 (0%)	0 (0%)
Test requirement can motivate hard work	0 (0%)	0 (0%)	6 (2%)

Table A15: Rationales behind Attitudes toward Test-Blind Admission Policies

Notes: This table shows the frequencies of spectators' rationales for their attitudes toward test-blind admission policies. Participants are categorized based on their reported support: "Support test-blind policies" includes those who somewhat or strongly support test-blind policies, "Support test-required policies" includes those who somewhat or strongly support test-required policies, and "Indifferent" refers to participants who are neutral between test-blind and test-required policies. The categories are summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

Reason	Support Test-Optional (N=239)	Indifferent (N=116)	Support Test-Required (N=238)
Mention usefulness of SAT/ACT scores	38 (16%)	11 (9%)	145 (61%)
Mention fairness of SAT/ACT scores	24 (10%)	6 (5%)	11 (5%)
Making SAT/ACT optional complicates decisions	1 (0%)	2 (2%)	8 (3%)
Making SAT/ACT optional biases decisions	8 (3%)	13 (11%)	17 (7%)
Test prep affects score validity	2 (1%)	0 (0%)	0 (0%)
Mention other determinants of score validity	49 (21%)	11 (9%)	113 (47%)
More information is better	5 (2%)	2 (2%)	10 (4%)
SAT/ACT could be down-weighted	7 (3%)	2 (2%)	9 (4%)
SAT/ACT could be viewed in context	3 (1%)	0 (0%)	0 (0%)
Students should be evaluated holistically	5 (2%)	2 (2%)	0 (0%)
SAT/ACT has a consistent standard	2 (1%)	0 (0%)	21 (9%)
Other admission criteria are more unfair	0 (0%)	0 (0%)	0 (0%)
Tests are stressful	5 (2%)	0 (0%)	0 (0%)
Test requirement can motivate hard work	0 (0%)	0 (0%)	1 (0%)
Mention freedom to submit/withhold scores	66 (28%)	4 (3%)	2 (1%)
Test-optional helps enhance applications	59 (25%)	4 (3%)	4 (17%)
Same criteria for all applicants	1 (0%)	3 (3%)	13 (5%)
Making SAT/ACT optional increases applications	9 (4%)	2 (2%)	1 (0%)

Table A16: Rationales behind Attitudes toward Test-Optional Admission Policies

Notes: This table shows the frequencies of spectators' rationales for their attitudes toward test-optional admission policies. Participants are categorized based on their reported support: "Support test-optional policies" includes those who somewhat or strongly support test-optional policies, "Support test-required policies" includes those who somewhat or strongly support test-required policies, and "Indifferent" refers to participants who are neutral between test-optional and test-required policies. The categories are summarized from spectators' open-ended responses. Each response can be assigned to one or more categories.

F Screenshots of the Survey

F.1 Admission Experiment Survey

Center for Behavioral and Decision Research

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?

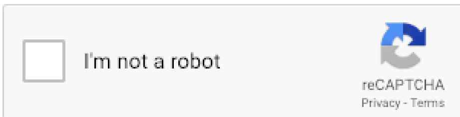
- ○ Very liberal
- ○ Somewhat liberal
- ○ Neither liberal nor conservative
- ○ Somewhat conservative
- ○ Very conservative
- ○ Other (please specify)
- ○ Prefer not to say

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!

F.2 Bonus Allocation Experiment Survey

Before continuing, please confirm that you are a human participant (i.e., not a "bot" program) by clicking on the checkbox below.



Prolific ID

What is your Prolific ID?
Please note that this response should auto-fill with the correct ID

Welcome message

Welcome! This study will take around **25 minutes**. So long as you complete the whole study, you will receive a **$6** completion payment. The decisions you make in this study may affect the payment of other Prolific workers, so please choose carefully.

Throughout the survey, we include questions that test your understanding of the instructions. You need to correctly answer these questions in order to proceed in the survey, so please read the instructions carefully.

Introduction to the context

Context

A group of workers were hired to collect information about companies online. The job had two 5-minute parts: Part A focused on financial companies, and Part B on service companies.

After finishing, each worker's submission was checked by an evaluator. The worker's score for each part is the number of companies they successfully recorded according to the evaluator:

For example, if someone got a score of 3 in Part A and 5 in Part B, it means the evaluator judged that they successfully recorded 3 financial companies and 5 service companies.

Context

The evaluator scored some workers on Day 1 and the rest on Day 2. (These two sets of workers do not differ in any significant way.) On Day 1, all workers were scored correctly for both parts. On Day 2, workers' Part A scores were also correct, but some workers' Part B scores were wrong. We will tell you more about the errors in Day 2 workers' Part B scores later.

Information from the previous screen:

A group of workers were hired to collect information about companies online. The job had two 5-minute parts: Part A focused on financial companies, and Part B on service companies.

After finishing, each worker's submission was checked by an evaluator. The worker's score for each part is the number of companies they successfully recorded according to the evaluator:

For example, if someone got a score of 3 in Part A and 5 in

Part B, it means the evaluator judged that they successfully recorded 3 financial companies and 5 service companies.

Your decisions

Every worker who completed the information collection task received the same base payment. In addition, some may receive an additional bonus of $3. In this survey, we will ask you to make decisions about how to allocate the bonus.

Specifically, you will make two decisions. Each decision will involve a pair of workers, and you will decide which one gets the bonus.

Your decisions

The bonus allocation will follow two steps:

First, We will generate a report card for each worker based on available information. The report card will always include the worker's Part A score and the day they were evaluated. It may also include their Part B score.

Next, You will choose a rule to decide which worker receives the bonus. The rule can use any information included on the report cards and may involve randomization (e.g., flipping a coin). However, it cannot depend on information not shown on the report cards.

Information from the previous screen:

Every worker who completed the information collection task received the same base payment. In addition, some may receive an additional bonus of $3. In this survey, we will ask you to make decisions about how to allocate the bonus.

Specifically, you will make two decisions. Each decision will involve a pair of workers, and you will decide which one gets the bonus.

Your decisions

One of your decisions may be randomly selected and used to determine a real bonus payment, so please consider each choice carefully. Whether they are implemented or not, the decisions will be completely anonymous. The workers will not know who made the decision, or even that their bonus was chosen by a third party.

Now, if you are ready, please click Next to proceed to the first bonus allocation decision.

Introduction to Accurate info part

Decision 1

For this decision, you will choose which of two workers receives a $3 bonus, **both workers were evaluated on Day 1.**

Accurate info part, Exclude Task 2 condition

First, please choose a bonus allocation rule assuming the report card includes only **1) each worker's Part A score and 2) the day they were evaluated**. The Part B score will *not* be included.

Recall that Part A scores are accurate and free of error.

Below is an example of what the two workers' report cards might look like:

Worker Scored on Day 1 Part A score: X	**Worker** Scored on Day 1 Part A score: Y

To make sure you understand the information on this screen, we'll ask some comprehension questions next. Click Next when you're ready.

On which days were the two workers' performances evaluated?

- ○ Both were evaluated on Day 1
- ○ Both were evaluated on Day 2
- ○ One was evaluated on Day 1 and the other on Day 2

Your answer was incorrect, please try again:

On which days were the two workers' performances evaluated?

- Both were evaluated on Day 1
- Both were evaluated on Day 2
- One was evaluated on Day 1 and the other on Day 2

Which of the following statements is correct?

- The two workers' Part A scores are accurate and free of error
- The two workers' Part A scores may be inaccurate

Your answer was incorrect, please try again:

Which of the following statements is correct?

- The two workers' Part A scores are accurate and free of error
- The two workers' Part A scores may be inaccurate

You passed the comprehension check.

Now, **which of the following rules would you prefer to**

use for allocating the bonus? *Your chosen rule may directly determine who receives the bonus.*

Flip a coin to decide who gets the bonus, regardless of their scores.	Give the bonus to the worker with the higher Part A score. If their Part A scores are the same: flip a coin to decide.
○	○

Accurate info part, Include Task 2 condition

Next, we ask you to choose a bonus allocation rule assuming the report card shows each worker's **1) Part A score, 2) Part B score, and 3) the day they were evaluated**.

Recall that since both workers were evaluated on Day 1, their scores for both parts are accurate—there are no errors.

Below is an example of what the two workers' report cards might look like:

Worker	Worker
Scored on Day 1	Scored on Day 1
Part A score: X	Part A score: Y
Part B score: Z	Part B score: W

To make sure you understand the information on this screen, we'll ask some comprehension questions next. Click Next when you're ready.

Which of the following scores for the two workers are accurate and free of error? (Check all that apply)

- ☐ The two workers' Part A scores
- ☐ The two workers' Part B scores

[If answer was incorrect]

Your answer were incorrect, please try again:

Which of the following scores for the two workers are accurate and free of error? (Check all that apply)

- ☐ The two workers' Part A scores
- ☐ The two workers' Part B scores

You passed the comprehension check.

Now, **which of the following rules would you prefer to use for allocating the bonus?** *Your chosen rule may directly determine who receives the bonus.*

Give the bonus to the worker with the higher Part A score. If their Part A scores are the same, flip a coin to decide.	Give the bonus to the worker with the higher total score (Part A + Part B). If their total scores are the same, flip a coin to decide.
○	○

Could you explain why you chose this bonus allocation rule?

Accurate info part, information preference

You've indicated your preferred rule:

- When Part B score is excluded from the report card:
  - [participant's chosen rule]

- When Part B score is included in the report card:
  - [participant's chosen rule]

Recall: Both workers were evaluated on Day 1, and their scores for both Part A and Part B are accurate.

Now, please let us know **whether you want the report cards to include Part B scores**. When we implement your bonus allocation decision, we will most likely make the report cards the way you prefer.

I want the report cards to **exclude** Part B scores	I'm indifferent	I want the report cards to **include** Part B scores
○	○	○

[If "I want the report cards to include Part B scores" is chosen]

To confirm your preference, please type the following sentence into the text box below.

I want the report cards to include Part B scores.

You can also choose not to type this sentence. In that case, we will interpret your preference as indifference.

[If "I want the report cards to exclude Part B scores" is chosen]

To confirm your preference, please type the following sentence into the text box below.

I want the report cards to exclude Part B scores.

You can also choose not to type this sentence. In that case, we will interpret your preference as indifference.

Could you explain why you have this preference?

Introduction to Invalid and Biased info part

Decision 2

For this decision, you will allocate the bonus between a new pair of workers. Unlike the previous decision, **one worker was evaluated on Day 1, and the other on Day 2.**

As in the previous decision, we will first generate a report card for each worker containing some information. Then, you will choose a bonus allocation rule that uses only the information shown on the report cards.

Invalid and Biased info part, Exclude Task 2 condition

First, please choose a bonus allocation rule assuming the report card includes only **1) each worker's Part A score and 2) the day they were evaluated**. The Part B score will *not* be included.

Recall that Part A scores are accurate and free of error.

Below is an example of what the two workers' report cards might look like:

Worker Scored on Day 1 Part A score: X	**Worker** Scored on Day 2 Part A score: Y

To make sure you understand the information on this screen, we'll ask some comprehension questions next. Click Next when you're ready.

On which days were the two workers' performances evaluated?

- ○ Both were evaluated on Day 1
- ○ Both were evaluated on Day 2
- ○ One was evaluated on Day 1 and the other on Day 2

[If answer was incorrect]

Your answer was incorrect, please try again:

On which days were the two workers' performances evaluated?

- ○ Both were evaluated on Day 1
- ○ Both were evaluated on Day 2
- ○ One was evaluated on Day 1 and the other on Day 2

Which of the following scores for the two workers are accurate and free of error? (Check all that apply)

- ☐ Day 1 worker's Part A score
- ☐ Day 2 worker's Part A score

[If answer was incorrect]

Your answer was incorrect, please try again:

Which of the following scores for the two workers are accurate and free of error? (Check all that apply)

- ☐ Day 1 worker's Part A score
- ☐ Day 2 worker's Part A score

You passed the comprehension check.

Now, **which of the following rules would you prefer to use for allocating the bonus?** *Your chosen rule may directly determine who receives the bonus.*

Give the bonus to the worker with the higher Part A score. If their Part A scores are the same: flip a coin to decide.	Flip a coin to decide who gets the bonus, regardless of their scores.
○	○

Invalid and Biased info part, Include Task 2 condition

Next, we ask you to choose a bonus allocation rule assuming the report card shows **1) each worker's Part A score, 2) Part B score, and 3) the day they were evaluated**.

The Day 2 worker's Part B score may have some errors in it. **Specifically, the evaluator incorrectly gave a random half of the Day 2 workers one extra point for Part B, but didn't indicate who got the extra points. This means that for any given Day 2 worker, there is a 50% chance that their Part B score is one point too high.**

The other three scores—the Day 1 worker's scores for both parts and the Day 2 worker's Part A score—are correct.

Below is an example of what the two workers' report cards might look like:

Worker
Scored on Day 1
Part A score: X
Part B score: Z

Worker
Scored on Day 2
Part A score: Y
Part B score: W*
*: May include one extra point added incorrectly

To make sure you understand the information on this screen, we'll ask some comprehension questions next. Click Next when you're ready.

Which of the following scores for the two workers are definitely accurate and free of error? (Check all that apply)

- [ ] Day 1 worker's Part A score
- [ ] Day 2 worker's Part A score
- [ ] Day 1 worker's Part B score
- [ ] Day 2 worker's Part B score

[If answer was incorrect]

Your answer were incorrect, please try again:

Which of the following scores for the two workers are definitely accurate and free of error? (Check all that apply)

- [ ] Day 1 worker's Part A score
- [ ] Day 2 worker's Part A score
- [ ] Day 1 worker's Part B score
- [ ] Day 2 worker's Part B score

What is the error in Day 2 worker's Part B score?

- ○ With a 50% chance, the evaluator may have given the Day 2 worker one extra point for Part B.
- ○ The evaluator always gave the Day 2 worker one extra point for Part B.

[If answer was incorrect]

Your answer was incorrect, please try again:

What is the error in Day 2 worker's Part B score?

- ○ With a 50% chance, the evaluator may have given the Day 2 worker one extra point for Part B.
- ○ The evaluator always gave the Day 2 worker one extra point for Part B.

You passed the comprehension check.

Now, **which of the following rules would you prefer to use for allocating the bonus?** *Your chosen rule may directly determine who receives the bonus.*

Give the bonus to the worker with the higher Part A score. If their Part A scores are the same, flip a coin to decide.	Give the bonus to the worker with the higher total score (Part A + Part B). If their total scores are the same, give the bonus to the Day 1 worker. *
○	○

**: Under this rule, the one extra point the Day 2 worker may have got only matters when the two workers have the same true total performance. When their true total performance is the same, the Day 2 worker gets the bonus only if they received the extra point. Otherwise, the Day 1 worker receives the bonus. If one worker has a higher true total performance, that worker always gets the bonus.*

Could you explain why you chose this bonus allocation rule?

Invalid and Biased info part, information preference

You’ve indicated your preferred rule:

- When Part B score is excluded from the report card:
  - [participant's chosen rule]
- When Part B score is included in the report card:
  - [participant's chosen rule]

Recall: Both workers’ Part A scores were accurate. However, the evaluator may have incorrectly given the Day 2 worker one extra point to the Part B score.

Now, please let us know **whether you want the report cards to include Part B scores**. When we implement your bonus allocation decision, we will most likely make the report cards the way you prefer.

I want the report cards to **exclude** Part B scores	I’m indifferent	I want the report cards to **include** Part B scores
○	○	○

[If "I want the report cards to include Part B scores" is chosen]

To confirm your preference, please type the following sentence into the text box below.

I want the report cards to include Part B scores.

You can also choose not to type this sentence. In that case, we will interpret your preference as indifference.

[If "I want the report cards to exclude Part B scores" is chosen]

To confirm your preference, please type the following sentence into the text box below.

I want the report cards to exclude Part B scores.

You can also choose not to type this sentence. In that case, we will interpret your preference as indifference.

Could you explain why you have this preference?

Demographics

Thank you for completing the main part of the survey. Before the end, please answer a few questions about yourself.

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

End of survey

If you have any suggestions or comments on this survey, 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!