Durham University Business School

Department of Economics

Merit, Identity, and Redistribution: Experimental Evidence on Affirmative Action

Ashwini Deshpande, Veronique Gille, Rajesh Ramachandran & Andis Sofianos

Working Paper No. 5, 2026

Durham Business School working papers series ISSN 1749-3641

Department Economics
Durham University Business School
Mill Hill Lane
Durham DH1 3LB, UK
Tel: +44 (0)191 3345200
https://www.durham.ac.uk/business/about/departments/economics/

Merit, Identity, and Redistribution: Experimental Evidence on Affirmative Action*

Ashwini Deshpande† Veronique Gille‡ Rajesh Ramachandran§
Andis Sofianos¶

May 22, 2026

Abstract

Affirmative action (AA) is highly controversial. This study asks whether opposition to AA reflects a genuine commitment to merit, or whether pro-merit discourse conceals deeper identity-based considerations. Specifically, does resistance vary with the social identity of those who benefit from the policy? Using an incentivized online experiment with university students in India, we compare perceptions and rewards toward participants selected through caste-based versus income-based affirmative action. Evaluators assess test-takers' competence and allocate monetary rewards under different selection rules. We find that beneficiaries from India's historically stigmatized caste groups (Scheduled Castes and Tribes, or SC-ST) are perceived as less competent than non-marginalized beneficiaries, even under income-based affirmative action. This suggests that responses to affirmative action vary systematically with the social identity of beneficiaries rather than reflecting a generalized aversion to preferential treatment alone. Yet these negative perceptions do not translate into corresponding material penalties: allocations toward caste-based beneficiaries are directionally compensatory, particularly among low-income beneficiaries, whereas allocations under income-based affirmative action align more closely with perceived competence. Overall, our findings reveal a duality in responses to affirmative action: identity-driven competence stigma can coexist with redistributive behavior that does not penalize marginalized beneficiaries.

JEL Codes: C90, J71, J78

Keywords: Affirmative Action, Reservations, Caste, Class, India, Experiment

*We gratefully acknowledge funding received from the Indian Council of Social Science Research (ICSSR) as a part of the EQUIP project on "Challenging Inequalities: An Indo-European Perspective" that made this paper possible. We thank Loukas Balafoutas, Kalyani Choudhuri, Lata Gangadharan, Srijita Ghosh, Pushkar Maitra, and Claire Rimbaud for useful comments. Versions of this paper were presented at Erasmus Rotterdam University, WZB Berlin Social Science Center, Queen's University Belfast, Queen Mary Workshop on Behavioural Economics, Asia-Pacific Economics Science Association (ESA) Meeting, Third Wuhan Cherry Blossom Workshop in Experimental Economics, Aix-Marseille Challineq Conference. The usual disclaimer applies.

†Ashoka University, India. Email: ashwini.deshpande@ashoka.edu.in

‡IRD, UMR LEDa, PSL, Université Paris-Dauphine, France. Email: veronique.gille@ird.fr

§Monash University Malaysia. Email: rajesh.ramachandran@monash.edu

¶Durham University, UK. Email: andis.sofianos@durham.ac.uk

1 Introduction

When affirmative action beneficiaries face negative perceptions or backlash, what drives these responses? Is it opposition to preferential treatment as a principle—a belief that any deviation from pure meritocracy is unfair regardless of who benefits? Or does stigma attach primarily to the social identity of beneficiaries, reflecting deeper biases against historically marginalized groups? More fundamentally, can identity-neutral redistributive criteria reduce such stigma, or do social identities continue to shape perceptions even when disadvantage is framed in ostensibly neutral terms? Understanding this distinction is critical for both the design of redistributive policies and for evaluating their real-world impact.

Two competing explanations exist in the literature and policy discourse. The first holds that opposition to affirmative action stems from a generalized aversion to preferential treatment (Sniderman and Carmines; 1997; Kluegel and Smith; 1986). Under this view, any policy that gives advantages based on group membership—whether race, caste, gender, or class—violates principles of individual merit and equal treatment (Bobocel et al.; 1998; Son Hing et al.; 2002; Reyna et al.; 2006). If this explanation is correct, we should observe similar negative attitudes toward all affirmative action beneficiaries, regardless of their social identity. The second explanation suggests that apparent meritocratic objections often mask identity-based bias (Kinder and Sears; 1981; Sears and Henry; 2003; Federico and Sidanius; 2002). Under this view, stigma is concentrated on beneficiaries from historically stigmatized groups (Bobo; 1998; Tarman and Sears; 2005), and reframing affirmative action around ostensibly identity-neutral criteria (such as income or class) would not reduce or eliminate negative perceptions (Sniderman and Carmines; 1997).

Distinguishing between these two explanations empirically is challenging. In most settings, affirmative action policies are defined by a single criterion—race in the United States, gender in corporate board quotas in Europe, caste in India's traditional reservation system. This makes it difficult to separate general opposition to preferential treatment from bias against specific beneficiary groups. What is needed is a context where affirmative action operates simultaneously along different dimensions—allowing us to compare attitudes when the policy criterion changes but the social identity of beneficiaries can be held constant, or vice versa.

India provides a unique opportunity for such an analysis. The country has maintained group-based affirmative action (known as "reservations") for Scheduled Castes and Scheduled Tribes (SC-ST)—historically marginalized groups subjected to severe discrimination and social exclusion—since independence in 1947. In 2019, India introduced an additional income-based quota for "Economically Weaker Sections" (EWS), explicitly designed to be caste-neutral and target economic disadvantage rather than social identity. SC-ST individuals are excluded from the EWS quota (they remain eligible to caste-based reservations).

We leverage this institutional context to design an incentivized experiment that disentangles policy-based from identity-based responses to affirmative action. University students in India—who have lived experience with both types of policies—evaluate real test-takers and allocate monetary rewards. Evaluators are randomly assigned to one of three conditions: (1) pure merit-based selection (control); (2) caste-based affirmative action with a reserved slot for SC-ST test-takers (matching India's historical policy); and (3) income-based affirmative action with a reserved slot for low-income test-takers (matching the 2019 reform). Critically, we introduce an experimental extension not permitted under current Indian law: we allow low-

income SC-ST individuals to appear as beneficiaries of income-based affirmative action. This design enables us to hold the policy rationale constant (economic disadvantage) while varying beneficiary identity (SC-ST versus General Category), directly testing whether economic framing alone can overcome identity-based stigma.

We measure two distinct responses: (1) *perceived competence*, captured through evaluators' guesses of test-takers' actual performance on a cognitive ability test; and (2) *reward allocations*, measured through real monetary transfers that evaluators distribute among test-takers. This design allows us to distinguish beliefs from behavior and to examine whether negative perceptions translate into material penalties. It is important to note that our objective is not to assess whether perceived competence and subsequent reward allocations correspond to the true underlying ability of affirmative action beneficiaries, either in our experimental setting or in the real world. Instead, we are interested in whether and how different types of affirmative action in the selection process trigger (or not) different beliefs and behavioral responses. A potential concern is that responses may reflect social desirability bias. The comparison of responses across treatments, which are symmetrically subject to such pressures, allows us to interpret the observed differences as reflecting treatment effects rather than differential social desirability. In addition, the online and fully anonymous nature of the experiment further limits the scope for socially desirable responding (Mummolo and Peterson; 2019).

Our results reveal a complex dual pattern that supports both explanations but in different ways. We find clear evidence of identity-driven stigma: in the control condition, SC–ST test-takers are perceived as less competent than their non-SC–ST counterparts. This pattern is further amplified under affirmative action. SC-ST individuals admitted through caste-based affirmative action are perceived as substantially less competent than comparable non-beneficiaries (0.19 standard deviations lower), with the effect driven primarily by high-income SC-ST beneficiaries (0.27 standard deviations lower). Most strikingly, when we compare individuals with identical eligibility under income-based affirmative action, General Category beneficiaries receive a strong positive competence appraisal (0.43 standard deviations above baseline), while low-income SC-ST receive a positive competence appraisal that is half the size (0.22 standard deviations lower). Even under an ostensibly identity-neutral policy, caste identity continues to shape perceptions of merit and competence.

Yet these negative perceptions do not translate into corresponding redistribution penalties. In the control condition, SC-ST and low-income test-takers tend to receive more generous allocations than their counterparts. This pattern appears somewhat stronger under caste-based affirmative action, where beneficiaries—particularly those from low-income households—often receive rewards exceeding what their perceived competence would predict. Under income-based affirmative action, by contrast, rewards align more closely with perceived competence.

These findings make three contributions. First, our design, grounded in a real institutional context, contrasts with earlier studies that rely on artificial group labels (e.g., colored badges or fictitious categories) to avoid confounds (Gille; 2022). For instance, Petters and Schroeder (2020) show that recipients of randomly assigned affirmative action (i.e., not based on true participant characteristics) are rated lower in performance, particularly by raters who themselves benefited from the policy. The closest designs to ours are those of Balafoutas et al. (2016) and Dorrough et al. (2016), who compare responses to gender-based quotas versus benefits assigned through artificial criteria. Balafoutas et al. (2016) find that the nature of the beneficiary group matters for performance evaluations, while Dorrough et al. (2016) find

no effect on cooperative behavior. By leveraging India's dual affirmative action system and introducing an experimental extension, our design allows us to disentangle the role of the eligibility criterion from that of beneficiary identity in shaping responses to affirmative action.

Additionnaly, in our experimental setting, evaluators' own payoffs are not affected by their allocation of rewards. Our paper, thus, makes a distinct contribution in contrast to the literature that identifies support or opposition to affirmative action based on self-serving motives (for instance, Herzog et al.; 2023).

Second, we jointly examine stigma and material responses within a unified framework. Previous research has documented both perceived incompetence (stigma) and material penalties (backlash) separately: women perceived to be hired through affirmative action are often viewed as less competent (Heilman et al.; 1992, 1997), with similar perceptions documented in the Indian context towards students admitted on caste-based AA (Deshpande; 2019); while other studies show that affirmative action can lead to backlash against women leaders selected through quotas (Gangadharan et al.; 2016), or to the systematic under-reporting of women's performance in competitive settings with peer evaluations (Leibbrandt et al.; 2018). Our approach allows us to show that stigma and backlash are distinct phenomena that need not move together.

Third, our findings speak to broader debates about the politics of redistribution and the role of identity versus class in shaping redistributive preferences. A well-established literature demonstrates that inequality is regarded as more acceptable when it is perceived to arise from merit or effort (e.g., Alesina and Angeletos; 2005; Almås et al.; 2010; Cappelen et al.; 2007, 2013; Fong; 2001; Konow; 2000). Recent work further shows that the perceived fairness of affirmative action depends on its rationale — whether it is instituted in response to discrimination, disadvantage, or historical injustice (Schildberg-Hörisch et al.; 2023). Our analysis is distinct in that we examine whether the same policy rationale is interpreted differently depending on the identity of the beneficiary. By comparing caste, income, and income-within-caste, we study how social identity and economic disadvantage interact in shaping perceptions of competence and deservingness. This has direct implications for policy debates on identity- versus class-based affirmative action. A related question is whether evaluators' own caste shapes their beliefs and reward allocations, for instance through in-group bias. Our experiment is not powered to examine this dimension, which we leave to future research.

The remainder of the paper proceeds as follows. Section 2 provides the institutional context of affirmative action in India. Section 3 describes the experimental design. Section 4 outlines the estimation framework. Section 5 presents the results. Section 6 discusses the findings and concludes.

2 Context and Institutional Setting

India's affirmative action (AA) program is complex in terms of its design and vast in terms of coverage. It is primarily quota-based, with positions reserved for eligible candidates in government-run higher educational institutions, government jobs, and electoral seats. Thus, it is called the *reservation policy*. In addition to reservation of positions, there are also financial measures such as fee concessions or priority credit for designated purposes that together make up the AA policy. Reservations in India are both "vertical" for mutually exclusive groups, i.e., individuals cannot simultaneously belong to more than one category (e.g., castes or tribes),

and "horizontal", targeting categories that are cross-cutting across the vertical categories (e.g., women or people with disabilities).

There are currently four groups eligible for vertical AA. The Scheduled Castes (SCs) is a group of jatis or castes that are at the bottom of the caste hierarchy, and were historically subjected to the highly degrading and stigmatising practice of untouchability. Untouchability has been legally abolished in independent India, and it is a criminal offense, but the stigma of belonging to a formerly untouchable jati runs deep, with covert and overt instances of untouchability continuing to be practiced (Thorat and Joshi; 2020). Distinct from the caste system, India is home to indigenous and tribal populations. Several of these are targeted for AA under the second vertical category called the Scheduled Tribes (STs). According to the reservation policy, 22.5% of seats in government-run educational institutions for students, teachers and administrative staff, as well as jobs in public sector enterprises and in government bureaucracy are reserved for the Scheduled Castes and Scheduled Tribes (SC-ST). The third group in the vertical reservations scheme is called *Other Backward Classes* (OBCs), which is a heterogeneous group of castes and communities that are, in principle, low in socioeconomic hierarchy but have not been subjected to the stigma of untouchability. An additional 27 percent positions are reserved for OBCs. We do not consider this group in our experiment.

In 2019, through a constitutional amendment, another category called *Economically Weaker Sections* (EWS) was added to the vertical reservations with an additional 10 percent quota. This was, in principle, meant to address class or income-based deprivation, distinct from the earlier group-based quotas. While the concept of EWS was meant to invoke the poor, in reality, the income cut-off set to identify the EWS (those with annual household incomes below INR 800,000) is set so high that over 95 percent of Indians would be eligible[1]. Additionally, by making this a vertical category, members belonging to SC, ST, or OBC categories are excluded from it, thereby effectively making it a quota for the so-called upper castes (Deshpande and Ramachandran; 2022). This was challenged in the courts, but was upheld by the Supreme Court in 2022.

India also passed "The Right of Children to Free and Compulsory Education (RTE) Act, 2009" for school children. Under this Act, private schools have to reserve seats for EWS children and provide education free of cost. The income limit for the EWS category for the purpose of RTE is INR 100,000 which captures economic deprivation much more accurately compared to the income cut-off for the quotas in higher education. In this paper, we have used the RTE EWS cut-off, since we want to assess attitudes regarding redistribution towards the economically deprived.

In addition, there is a parallel component of the affirmative action programme in the electoral sphere, which consists of seats reserved for SC-STs at all electoral levels, and for women at the local body levels.

[1]There are very few reliable sources of data for income distribution. Our estimates from the unit-level data of the Consumer Pyramids Household Survey (CPHS), gathered by Centre for Monitoring Indian Economy (CMIE), reveals that almost 98 percent of households have an income below INR 800,000

3 Experimental Design

We implement an online experiment which consists of two parts. The first part involves *test-takers* completing a cognitive ability test. In the second part, a separate set of participants, college and university students referred to as *evaluators*, assess the performance of groups of test-takers and allocate a bonus reward among them. Depending on the treatment, selected test-takers for the bonus are either the top scorers in their group or a combination of top scorers and affirmative action beneficiaries. Table 1 summarizes the sequence of the experiment and the selection rules across treatment arms; these aspects are discussed in detail below.

Table 1: Experimental Design and Study Flow

<table>
<tr><th>Stage</th><th colspan="4">Description</th></tr>
<tr><td>Test-takers</td><td colspan="4">Participants complete the Raven Advanced Progressive Matrices (APM) test and a demographic questionnaire.</td></tr>
<tr><td>Group formation and selection</td><td colspan="4">Groups of five are formed from the pool of test takers with variation in caste and income composition. Within each group, four test-takers are selected for possible reward allocation according to the rules below.</td></tr>
<tr><td></td><td>Control
No AA</td><td>AA–Caste
Caste-based AA</td><td colspan="2">AA–Income
Income-based AA</td></tr>
<tr><td></td><td></td><td>Institutional</td><td>Institutional</td><td>Extended</td></tr>
<tr><td></td><td>Top four overall selected</td><td>Top three overall selected, plus one reserved slot for an SC–ST candidate</td><td>Top three overall selected, plus one reserved slot for a low-income GC candidate</td><td>Top three overall selected, plus one reserved slot for a low-income GC or SC–ST candidate</td></tr>
<tr><td>Evaluator tasks</td><td colspan="4">Evaluators first guess selected test-takers' Raven scores. The total reward pot is then generated from the actual scores of the selected test-takers, after which evaluators allocate the pot among them.</td></tr>
<tr><td>Post-experiment</td><td colspan="4">Evaluators complete the Cognitive Reflection Test (CRT), demographic questionnaire, and debriefing questions.</td></tr>
</table>

Notes: The table summarizes the sequence of the experiment and the selection rules across treatment arms. In the control condition, the four highest scorers in each group are selected. Under the institutional caste-based rule, the three highest scorers are selected together with one reserved slot for an SC–ST candidate. Under the institutional income-based rule, the reserved slot is assigned to a low-income General Category (GC), i.e. non-SC–ST, test-taker. Under the extended income-based rule, the reserved slot is open to any low-income test-taker, including SC–ST individuals.

3.1 Test-taker Tasks

As Table 1 shows, the first stage of the experiment involves test-takers completing the Raven Advanced Progressive Matrices (APM) test and a demographic questionnaire. Specifically,

they are asked to complete a short version of the Raven Advanced Progressive Matrices (APM) (Raven; 1962). This is a non-verbal test which abstracts from any language or math skills. It is widely used to measure fluid intelligence and has received increased attention in the recent economics literature (some examples: Basteck and Mantovani; 2018; Carpenter et al.; 2013; Charness et al.; 2018; Gill and Prowse; 2016; Proto et al.; 2019, 2022). Given that we implement the experiment online, shortening the duration of the test was important.[2] Hence, we follow Bors and Stokes (1998) in using 12 out of a total of 36 matrices from Set II of the APM. This specific test is typically used to elicit the ability of above average intelligence individuals, such that it is objectively a cognitively difficult test. The participants are allowed a maximum of eight minutes to complete this test. For each question, a 3×3 matrix of images is displayed on the participants' screen with the image in the bottom right corner missing. The participants are then asked to complete the pattern choosing one out of eight possible choices presented on the screen. The 12 matrices are presented in increasing order of difficulty exactly as sequenced in Set II of the APM. During the eight minutes, participants are allowed to switch back and forth across the 12 matrices and change their answers. For each correct answer, participants are rewarded with INR 25.

Once participants complete the Raven test they are asked to respond to a demographic questionnaire about their age, gender, caste, and household income. Descriptive statistics for the sample of participants are reported in Table A1 in Appendix. The complete questionnaire is included in the Appendix (see figures A5 and A6 in Appendix A6.1). There was no payment for completing this part.

After completing the questionnaire, participants are informed about how much they have earned including a show-up fee of INR 200. On this last screen, participants are also informed that they may potentially be selected to receive an additional bonus in the coming weeks (see Figure A7 in Appendix A6.1 for the exact phrasing).

3.2 Group Formation and Affirmative Action Treatments

From the set of all test-takers, we form groups of five. We always ensure that each group of five includes GC and SC-ST individuals, as well as test-takers that are from both low and high-income households. Within each group, four group members are selected to receive a bonus reward. Selection is either based exclusively on Raven scores, or one position is reserved for an SC-ST or low-income individual. The embedded box in Table 1 summarizes the selection rules for the different treatment arms.

Crucially, the experiment is designed such that test-takers have identical characteristics across the control and treatment conditions. What varies across conditions is the selection rule, and therefore which test-takers are chosen to receive a reward. In the control arm (no AA), the four highest-scoring test-takers out of each group of five are selected to receive the bonus reward. In this scenario there are no reserved positions.

In the caste-based AA treatment (AA-Caste), three individuals are selected based on attaining the top three test scores. The fourth position reflects the reservation policy first established in 1950: the reserved slot is allotted to a SC-ST participant. If both of the

[2]We implemented the experiment in late 2022 and early 2023, before the current implosion of AI tools. We can safely assume that the participants would not have used AI to solve the puzzles.

remaining participants are SC-ST, the higher scorer receives the slot, while if only one is SC-ST, that individual is automatically selected.

In the income-based AA treatment (AA-Income), which parallels the more recent extension of quotas to the Economically Weaker Sections (EWS), the individuals with the top three test scores are again selected, while the reserved slot is allotted to a low-income candidate, i.e. to individuals with annual household income below INR 100,000.

Within the income-based AA treatment, the *Institutional* selection rule mimics the affirmative action policies embedded in India's institutional framework: the reserved slot is allotted to a low-income General Category (GC).[3] The *Extended* selection rule modifies the institutional arrangement by reserving the last slot for all low-income participants, including those from the SC-ST category.

Extended setting. This introduces a novel experimental comparison by allowing SC-ST individuals to also benefit from income-based affirmative action. This allows us to examine whether competence perceptions continue to vary across beneficiary groups even within an ostensibly income-based redistributive framework. In other words, the design tests whether responses to affirmative action depend on the identity of those who benefit from the policy.

3.3 Evaluator Tasks

As shown in Table 1, the third stage of the experiment involves evaluators, whose behavior is the focus of the experiment. They evaluate the competence and then allocate rewards to the test takers from stage 1. Evaluators gave their assessments online.

Evaluation of competence. Evaluators are first asked to guess the scores of four different groups of test-takers, which are formed as described in Section 3.2. Depending on the treatment arm, evaluators are informed about the criteria used to select recipients from each group for the reward (as described in Section 3.2). When making their guesses, evaluators are shown a grade corresponding to a range of scores (see Table 2) and asked to provide their best estimate of the actual score.[4] If a guess falls outside the range associated with a grade, the evaluator is alerted and asked to revise it. Score guesses are incentivized using the binarised scoring rule (Hossain and Okui; 2013), and following Danz et al. (2022), we emphasize that accurate reporting yields the highest expected payoff. One of the four groups is randomly selected, and for that group the evaluator can earn a bonus of INR 35 for each of the five guesses, calculated with the binarised scoring rule. Evaluators can access additional details of the payment procedure through supplementary information available by clicking a button on the screen.[5] Evaluators also receive information on each test-taker's caste, whether

[3]Since these are the non-SC-ST-OBC individuals, among Hindus they serve as a proxy for higher-ranked or so-called upper castes.

[4]Evaluators see test-takers listed in descending order of grades. This design mechanically implies that guessed scores should decline with position, which serves as a basic check of task comprehension. In analyzing competence evaluations, we therefore focus on treatment differences by comparing guessed scores for participants in the fourth position across the control and AA treatments. The mechanical decline within groups is, thus, not a concern for our results.

[5]See Figures A10, A11, and A12 in Appendix A6.2.

their household income is below INR 100,000, and — when relevant in the affirmative action treatments — their affirmative action status.[6]

Table 2: Evaluators: Grade & Percentage Score Correspondence

Grade	A	B	C	D
Score Range	$75-100$	$50-74$	$25-49$	$0-24$

Notes: The table shows the correspondence between grades and percentage score ranges provided to evaluators. Grades were used to summarize test-takers' performance, and evaluators entered point estimates within the corresponding range.

Reward allocation. The second task requires evaluators to allocate bonuses among the four individuals selected from each group of five test-takers. As Table 1 shows, this corresponds to the "Evaluator tasks" stage. Evaluators are informed that the total pot available for distribution equals the sum of the percentage scores of the selected individuals. Since evaluators have previously guessed the scores of these individuals, this information allows them to infer how accurate their beliefs were at the aggregate level. However, because they only observe the total sum of scores, rather than the individual scores themselves, they cannot infer whether their guesses about any particular individual were correct or incorrect. For example, if the selected participants scored $74, 68, 52$, and 60, the pot is 254 Experimental Currency Units (ECU). This rule is designed to make salient the potential perception of efficiency loss arising from affirmative action. For each ECU allocated, test-takers earn INR 0.1 (e.g., 100 ECUs = INR 10). Importantly, evaluators' own payments are unaffected by their allocations, but their decisions determine the additional bonus received by the selected test-takers.[7] While allocating rewards, evaluators are again shown each test-taker's grade, caste, and whether their household income is below INR 100,000. In the AA treatments, they are additionally informed of each test-taker's affirmative action status.

Other measures. After completing the allocation task, evaluators are asked to complete a cognitive reflection test (CRT).[8] Additionally, evaluators are asked to respond to a demographics questionnaire.[9] After evaluators are informed about their earnings, we also include a set of de-briefing questions.[10]

3.4 Implementation

The test-takers were recruited through Amazon Mechanical Turk (MTurk), and in total 57 participated, with an average payment of INR 332.30. The main experiment was administered online using oTree (Chen et al.; 2016). Evaluators were recruited from the college

[6]See Figures A19 - A30 in Appendix A6.2 for the screen interface for each of the four groups across all three treatments.

[7]Evaluators are not told how many other participants are also making allocation decisions.

[8]This is a test that was introduced by Frederick (2005) and evaluates the extent to which an individual is more intuitive or reflective. The questions are shown in figure A44 in appendix A6.2. We do not exploit the answers to this test in this paper.

[9]The full questionnaire is in figures A45, A46, and A47 in appendix A6.2.

[10]See figure A49 in appendix A6.2.

student population at the University of Delhi and from colleges in Mumbai and Pune.[11] Data were collected in two rounds: Round 1 in December 2022 with 201 evaluators, and Round 2 in April 2023 with 90 evaluators. The average payment for evaluators was INR 414.40. To ensure that evaluators understood the selection process for allocating bonus rewards, we administered a short comprehension quiz (see Figures A13–A18 in Appendix A6.2). In addition, we included checks throughout the study to verify that evaluators were attentive. The full set of screens and instructions is provided in Appendix A6. The design was pre-registered at the AEA RCT Registry (`https://www.socialscienceregistry.org/trials/10410`). While the planned sample size was 440 evaluators, recruitment proved challenging which meant we have a smaller sample than planned. This may limit power for some comparisons; however, the main estimates are statistically significant and economically meaningful. Ethics approval was obtained from the Ashoka University Institutional Review Board (22-X-10008-Deshpande).

3.5 Descriptive statistics

Table 3 presents descriptive statistics for the 291 evaluators included in our experiment, disaggregated by treatment status: control, AA-Caste, and AA-Income. The table reports mean values, with standard deviations shown in parentheses. Columns 4 and 5 test for statistically significant differences across treatment arms. The results indicate that evaluators' characteristics are balanced across conditions, which is expected given the random assignment. In the control group, the average evaluator is approximately 21 years old, consistent with the typical age of an undergraduate student in India. Around 41 percent are women, and approximately 79 percent identify as Hindu. Roughly 69 percent belong to the general caste category, about 18 percent come from households with annual incomes below INR 100,000, and 68 percent are enrolled in economics majors.

[11] The University of Delhi is one of the largest universities in India, with over 700,000 undergraduate students (https://www.du.ac.in/uploads/new-web/21082024˙AR˙Part-I-English.pdf). Mumbai and Pune also have a large population of undergraduate students in both public and private institutions. Mumbai and Delhi are two of the largest metropolitan cities in India, representing a microcosm of the Indian population.

Table 3: Comparison of evaluators' demographic characteristics across treatments

	Control	AA-Caste	AA-Income	(2)-(1)	(3)-(1)
Age	20.75	20.72	20.73	0.03	0.03
	(4.05)	(2.17)	(2.39)	(0.46)	(0.48)
Female	0.41	0.48	0.40	-0.07	0.01
	(0.49)	(0.50)	(0.49)	(0.07)	(0.07)
Hindu	0.79	0.77	0.76	0.02	0.03
	(0.41)	(0.42)	(0.43)	(0.06)	(0.06)
General Category	0.69	0.69	0.75	-0.00	-0.06
	(0.47)	(0.46)	(0.44)	(0.07)	(0.06)
OBC	0.22	0.21	0.16	0.01	0.05
	(0.41)	(0.41)	(0.37)	(0.06)	(0.06)
SC/ST	0.10	0.10	0.09	-0.01	0.01
	(0.30)	(0.31)	(0.28)	(0.04)	(0.04)
Income < INR 100 000	0.18	0.23	0.23	-0.05	-0.05
	(0.38)	(0.42)	(0.42)	(0.06)	(0.06)
Economics Major	0.68	0.73	0.70	-0.06	-0.02
	(0.47)	(0.45)	(0.46)	(0.06)	(0.07)
N observations	102	97	92		

Notes: This table reports mean demographic characteristics of evaluators across the three treatment groups. Standard deviations are in parentheses. Columns (4) and (5) show mean differences relative to the control group. None of the observed differences are statistically significant, indicating that random assignment generated balanced samples across treatments. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1 presents the average guessed scores (Panel A) and reward allocations (Panel B) for participants benefiting from affirmative action and for comparable participants in the same position without affirmative action in the control condition. At the aggregate level, the averages look very similar across treatment conditions, with only small differences. As we show in the regression analysis below, however, these similarities in the averages mask important patterns that emerge once we account for the income and caste identities of beneficiaries.

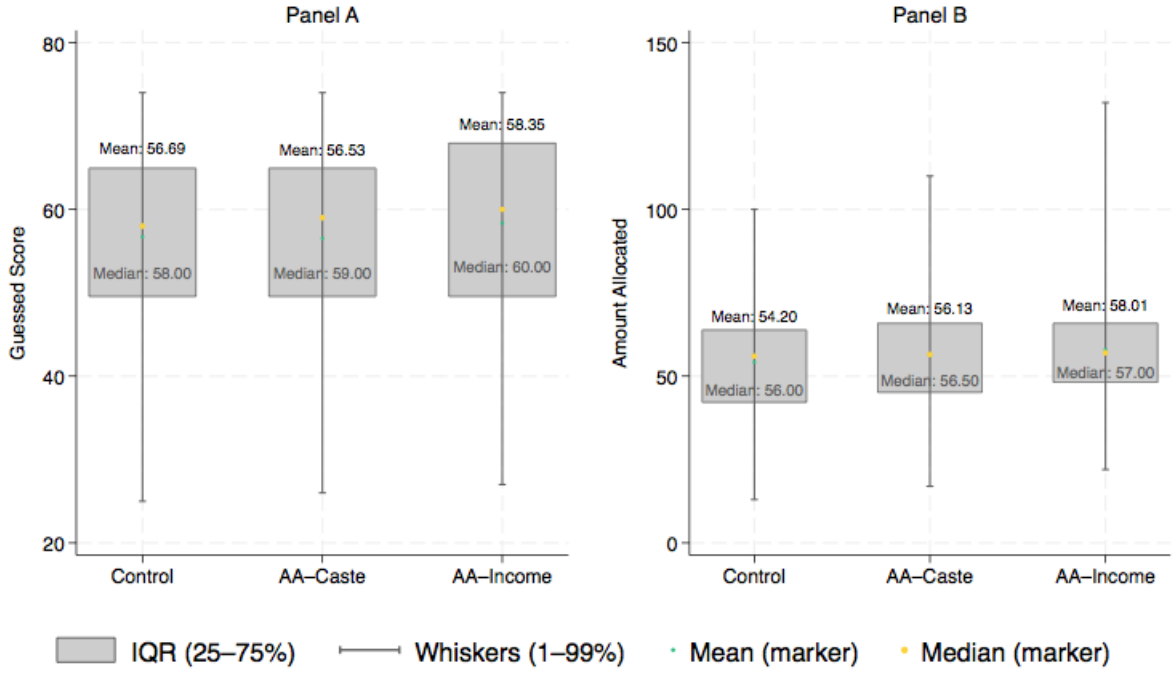


Figure 1: Boxplot of guessed scores and share allocated for participants in the fourth position

Notes: Figure 1 plots the mean, median, the inter-quartile range (25–75%), and the whiskers (1st and 99th percentile) of the guessed score and amount allocated under the control and two treatments for participants benefiting from affirmative action and for comparable participants in the same position without affirmative action in the control condition, i.e. in the fourth position.

4 Estimation framework

4.1 Control group

We begin by estimating equation (1) for the control group only. This provides a benchmark for how evaluators assess test-takers in the absence of any affirmative action intervention, allowing us to gauge whether identity-based rewards or penalties arise under the status quo. In particular, our coefficients of interest are α_1 and α_2, which capture the associations between test-taker caste status (SC-ST$_t$) and low-income status (Low-Income$_t$), respectively, and the outcomes of interest.

$$Y_{e,t,g} = \alpha_0 + \alpha_1 \text{SC}_t + \alpha_2 \text{Low-Income}_t + \alpha_3 \text{Position}_t + \alpha_4 \text{Grade}_t + \alpha_5 X_e + \alpha_6 \text{Round}_e + \alpha_7 \text{EvalOrder}_g + \epsilon_{e,t,g} \quad (1)$$

The outcome variable $Y_{e,t,g}$ is standardized to have mean zero and unit variance. We consider three distinct outcomes. First, perceived competence, measured by the score guessed by evaluator e for test-taker t. Second, the share of the total reward allocated to test-taker

t by evaluator e. Third, the deviation from score-based contributions, defined as the difference between the reward share allocated and the share implied by the evaluator's own score guesses for each test-taker. This third measure allows us to test whether evaluators allocate systematically more or less than what their own assessments of competence would justify.

We also control for Position_t, which denotes the position of the test-taker within the group; Test-taker's grade (Grade_t) captures the objective performance of the test-taker; and X_e is a vector of evaluator characteristics, including socio-demographic attributes. Round_e is a dummy for the round of data collection, and EvalOrder_g consists of indicator variables for the order in which groups are evaluated (group 1 through 4), capturing potential order effects in evaluators' assessments.

4.2 Institutional Setting

For the main specification, we focus on the test-taker presented in the fourth position. In the AA-treatment arms, this position is explicitly signaled as benefiting from affirmative action. The random allocation of evaluators to treatment arms implies the following specification, where t denotes the test-taker and e denotes the evaluator:

$$\begin{aligned} Y_{e,t,g} = \beta_0 &+ \beta_1 \text{AA-Treatment}_e + \beta_2 \text{Low-Income}_t + \beta_3 X_e \\ &+ \beta_4 \text{Round}_e + \beta_5 \text{EvalOrder}_g + \epsilon_{e,t,g} \end{aligned} \tag{2a}$$

$$\begin{aligned} Y_{e,t,g} = \gamma_0 &+ \gamma_1 \text{AA-Caste}_e + \gamma_2 \text{AA-Income}_e + \gamma_3 \text{Low-Income}_t \\ &+ \gamma_4 X_e + \gamma_5 \text{Round}_e + \gamma_6 \text{EvalOrder}_g + \epsilon_{e,t,g} \end{aligned} \tag{2b}$$

$$\begin{aligned} Y_{e,t,g} = \delta_0 &+ \delta_1 \text{AA-Caste}_e + \delta_2 (\text{AA-Caste}_e \times \text{Low-Income}_t) \\ &+ \delta_3 \text{AA-Income}_e + \delta_4 \text{Low-Income}_t + \delta_5 X_e \\ &+ \delta_6 \text{Round}_e + \delta_7 \text{EvalOrder}_g + \epsilon_{e,t,g} \end{aligned} \tag{2c}$$

The outcome variables are defined as in equation (1). AA-Treatment_e is a dummy equal to 1 if the evaluator is assigned to an affirmative action treatment. All other terms are defined as in equation (1).

Equation (2a) provides the most aggregated specification, where β_1 captures the effect associated with being assigned to any affirmative action treatment. Equation (2b) disaggregates this effect by decomposing AA-Treatment into caste- and income-based treatments, AA-Caste and AA-Income, with γ_1 and γ_2 capturing the effects associated with caste-based and income-based affirmative action, respectively.

Finally, equation (2c) further extends this framework by allowing the effect of caste-based affirmative action to vary with the income status of the beneficiary. Here, δ_1 captures the effect associated with caste-based affirmative action, δ_2 captures the additional effect for low-income test-takers under caste-based affirmative action, and δ_3 captures the effect associated with income-based affirmative action.

4.3 Extended Setting

Equation (3) further extends the framework by allowing both caste- and income-based affirmative action to vary with the underlying characteristics of the beneficiary. The coefficient ζ_1

captures the effect associated with caste-based affirmative action for non-low-income beneficiaries, while ζ_2 captures the additional effect for low-income beneficiaries under caste-based affirmative action. The coefficient ζ_3 captures the effect associated with income-based affirmative action for non-SC-ST beneficiaries, and ζ_4 captures the differential effect for SC-ST beneficiaries under income-based affirmative action. Finally, ζ_5 captures baseline differences associated with low-income status.

$$Y_{e,t,g} = \zeta_0 + \zeta_1 \text{AA-Caste}_e + \zeta_2 \text{AA-Caste}_e \times \text{Low-Income}_t + \zeta_3 \text{AA-Income}_e + \zeta_4 \text{AA-Income}_e \times \text{SC-ST}_t + \zeta_5 \text{Low-Income}_t + \zeta_6 X_e + \zeta_7 \text{Round}_e + \zeta_8 \text{EvalOrder}_g + \epsilon_{e,t,g} \quad (3)$$

5 Results

5.1 Control Group: Identity Effects

Figure 2 presents the estimated coefficients on the SC-ST and low-income indicators from equation 1, restricting the sample to the control condition. This control-group analysis provides a benchmark for how evaluators respond to caste and income identities in the absence of any affirmative action intervention. Coefficients on the remaining variables are reported in Appendix Table A2.

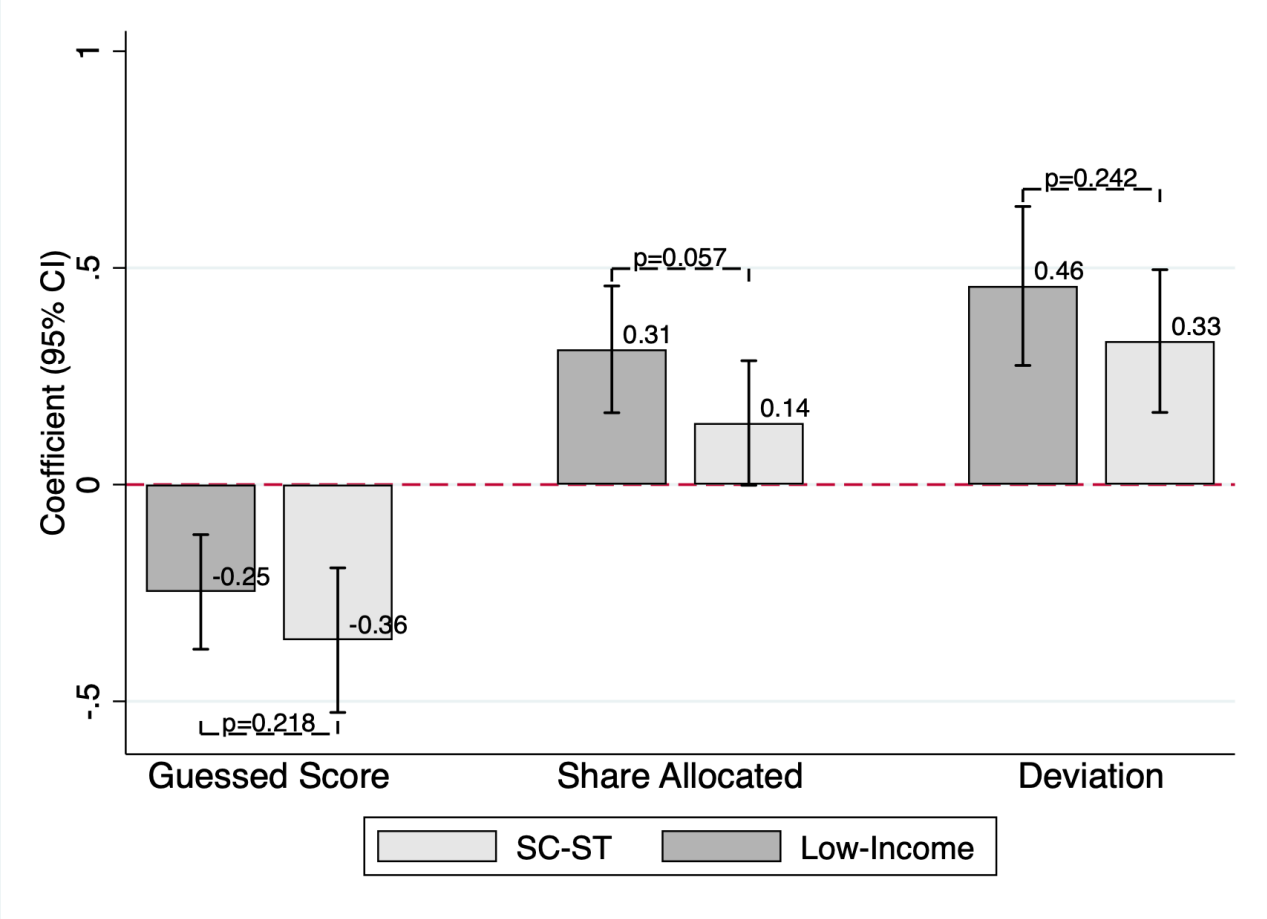
5.1.1 Guessed scores

Holding constant position, grade, and the order in which groups are evaluated, both low-income and SC-ST individuals are rated as less competent in the control group. Compared to high-income individuals, the guessed score for low-income test-takers is 0.25 SD lower, while compared to General Category individuals, the guessed score for SC-ST test-takers is 0.36 SD lower. Although both effects are individually significant, the equality of the two coefficients cannot be rejected ($p = 0.218$), indicating that income- and caste-based competence penalties are of broadly comparable magnitude in the absence of any affirmative action intervention.

5.1.2 Share allocated

A different pattern emerges for reward allocations. Holding all else constant, low-income individuals receive 0.31 SD more of the allocated share than their high-income counterparts, and SC-ST individuals receive 0.14 SD more than General Category individuals. The difference between these two effects is statistically significant at the 10% level, indicating that evaluators compensate economic disadvantage somewhat more generously than caste-based disadvantage, even though both groups are favored relative to their reference categories.

Figure 2: Baseline Identity Effects in the Control Condition



Notes: The figure reports estimated marginal effects of test-taker caste and income status on the three outcomes of interest in the control condition. Regressions control for test-taker position, grade, group evaluation order, and evaluator demographic characteristics. Vertical bars report 95% confidence intervals. P-values report tests of equality between the SC-ST and low-income coefficients for each outcome. Full regression results are reported in Appendix Table A2.

5.1.3 Deviation (share allocated minus share implied by guessed scores)

The deviation measure—defined as the share allocated minus the share implied by the evaluator's own guessed scores—reinforces this pattern. Low-income individuals are allocated 0.46 SD more than high-income individuals, and SC-ST individuals are allocated 0.33 SD more than General Category individuals. Both coefficients are statistically significant at the 5% level, but the equality of the two cannot be rejected.

Taken together, all six coefficients across the three outcomes are individually significant at the 5% level, while between-group differences are only significant for the share allocated.

5.2 Affirmative Action, Merit and Redistribution

This section examines how affirmative action affects evaluators' perceptions of competence and subsequent reward allocations. We first study how affirmative action shapes guessed scores,

which capture perceived competence, before turning to reward allocations and deviations between allocations and evaluators' own competence assessments. Together, these outcomes allow us to distinguish between competence stigma and redistributive behavior toward affirmative action beneficiaries.

5.2.1 Perceptions of competence

Table 4 reports estimates of the affirmative action treatment effects from equations 2a–3, where the dependent variable is the evaluator's guessed score for the test-taker in the fourth position. Columns (1)–(4) present estimates from the institutional and extended specifications, while Figure 3 summarizes the implied treatment effects by caste and income group under the extended setting.

Table 4: Impact of AA treatment on guessed score

	Outcome: Guessed Score			
	Institutional			Extended
	(1)	(2)	(3)	(4)
Low-Income	0.129	-0.207	-0.237	-0.337**
	(0.122)	(0.129)	(0.151)	(0.135)
AA treatment	0.106			
	(0.0853)			
AA-Caste		-0.192*	-0.238*	-0.268**
		(0.105)	(0.138)	(0.131)
AA-Caste × Low-Income			0.0853	0.295**
			(0.159)	(0.143)
AA-Income		0.299***	0.307***	0.433***
		(0.102)	(0.106)	(0.113)
AA-Income × SC-ST				-0.208*
				(0.115)
Constant	-0.354	0.00794	0.0226	0.124
	(0.337)	(0.330)	(0.332)	(0.310)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	291	291	291	291
N Observations	970	970	970	1164

Notes: The dependent variable is the evaluator's guessed score for the test-taker in the fourth position. Raw guessed scores are rescaled to a common 0–25 scale across grade levels and then standardized (z-scores). Column (1) corresponds to the baseline specification in Equation (2a), Column (2) to the institutional specification in Equation (2b), Column (3) to the institutional specification with income interactions in Equation (2c), and Column (4) to the extended specification in Equation (3). All regressions include evaluator socio-demographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Institutional Setting. Column (1) shows no statistically significant average effect of affirmative action on perceived competence. However, this aggregate pattern masks substantial heterogeneity across institutional settings. In Column (2), caste-based affirmative action beneficiaries are perceived as less competent, with guessed scores falling by 0.192 SD relative to comparable individuals in the control group. By contrast, beneficiaries under income-based affirmative action receive a positive and statistically significant competence appraisal of 0.299 SD.

Column (3) shows that the negative perception attached to caste-based affirmative action does not significantly differ between low-income and non-low-income SC-ST beneficiaries. Taken together, Columns (1)–(3) indicate that evaluators respond differently across the institutional affirmative action regimes, with caste-based beneficiaries perceived less favorably than beneficiaries under income-based affirmative action.

Extended Setting. While Columns (1)–(3) compare perceptions across the institutional affirmative action regimes, the extended setting in Column (4) allows us to examine whether competence perceptions continue to vary across beneficiary groups even within an ostensibly income-based redistributive framework.

Column (4) in Table 4 presents the estimated coefficients, and Figure 3 summarizes the implied treatment effects by group. Turning first to caste-based affirmative action, Figure 3 shows that negative competence perceptions are concentrated among non-low-income SC-ST beneficiaries, whose guessed scores are 0.27 SD lower than those of SC-ST individuals in the control group. By contrast, low-income SC-ST beneficiaries face no additional competence penalty under caste-based affirmative action; the estimated effect is close to zero.

Under income-based affirmative action, low-income General Category beneficiaries receive a strong and statistically significant positive competence appraisal of 0.43 SD relative to SC-ST individuals in the control group. Low-income SC-ST beneficiaries under the same policy also receive a positive appraisal, but its magnitude (0.22 SD) is substantially smaller, and the difference between the two estimates is marginally statistically significant ($p = 0.071$). Thus, even within an income-based affirmative action framework, competence perceptions continue to vary systematically across beneficiary groups.[12]

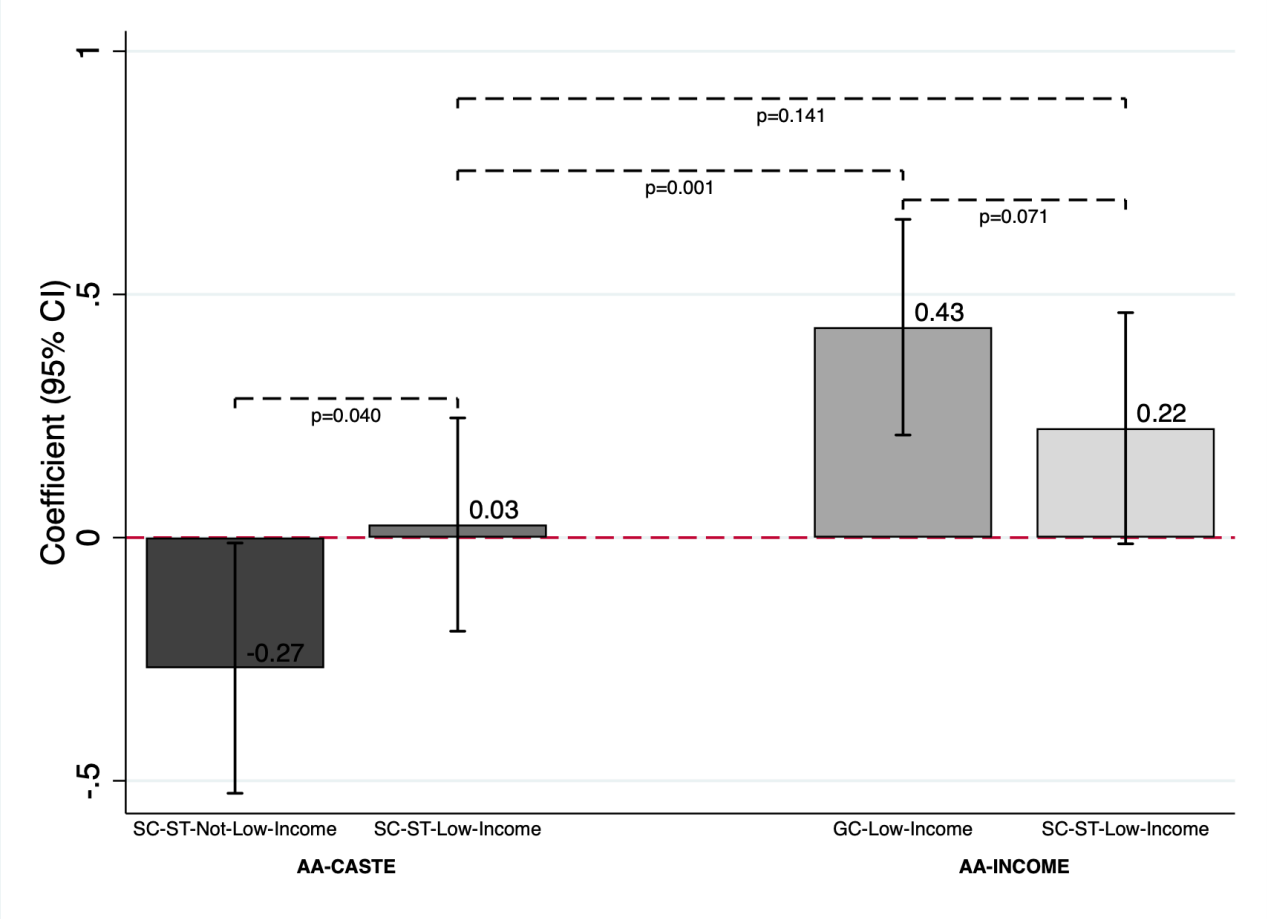
5.2.2 Redistribution responses

While Section 5.2.1 documented substantial heterogeneity in perceived competence across affirmative action beneficiaries, it remains unclear whether these perceptions translate into redistributive behavior. We therefore next examine how evaluators allocate rewards across beneficiaries, and whether these allocations align with evaluators' own assessments of competence.

Reward allocation. Columns (1)–(4) in Table 5 report the estimated coefficients from equations 2a–3, where the outcome is the share of the reward allocated to the test-taker in

[12]Comparing across treatments, the effect of income-based affirmative action on low-income SC-ST individuals (+0.22 SD) is larger than the corresponding effect of caste-based affirmative action (+0.03 SD), although the two estimates are not statistically distinguishable ($p = 0.141$).

Figure 3: Impact of AA treatment on guessed score, by income and caste group



Notes. Marginal effects of affirmative action treatments on guessed scores of beneficiaries, by income and caste groups. Marginal effects are computed using the specification in Column (4) of Table 4.

the fourth position. Figure 4 summarizes the implied treatment effects by caste and income group under the extended setting.

Institutional setting. The allocation results differ markedly from the competence perceptions documented in Section 5.2.1. Column (1) of Table 5 shows that affirmative action beneficiaries receive 0.24 SD more of the total reward than comparable non-beneficiaries. Unlike the guessed-score results, however, this allocation premium does not differ significantly across the two institutional affirmative action regimes. Column (2) shows that estimated allocation responses are directionally similar under both caste- and income-based affirmative action, despite substantial differences in perceived competence across the two groups; the coefficient on AA-Income is however statistically insignificant (p=0.13). Column (3) further indicates that allocations under caste-based affirmative action do not vary significantly by income status.

Extended setting. Panel A of Figure 4 and Column (4) of Table 5 extend the analysis by allowing SC-ST individuals to also qualify under income-based affirmative action. In contrast

Table 5: Impact of AA treatment on reward allocation

Outcome:	Share Allocation				Deviation from guessed score			
	Institutional			Extended	Institutional			Extended
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Low-Income	0.0699	0.0716	0.0517	0.193	-0.000192	0.113	0.135	0.331**
	(0.109)	(0.116)	(0.123)	(0.124)	(0.125)	(0.128)	(0.142)	(0.146)
AA treatment	0.240*				0.236*			
	(0.123)				(0.128)			
AA-Caste		0.242**	0.211	0.131		0.337***	0.370**	0.278*
		(0.118)	(0.138)	(0.130)		(0.124)	(0.154)	(0.145)
AA-CASTE × Low-Inc.			0.0566	0.139			-0.0609	-0.00221
			(0.145)	(0.140)			(0.168)	(0.164)
AA-Income		0.239	0.244	0.180		0.171	0.166	0.0608
		(0.159)	(0.159)	(0.174)		(0.166)	(0.167)	(0.188)
AA-Income × SC-ST				0.0509				0.108
				(0.112)				(0.121)
Constant	-0.468	-0.470	-0.460	-0.623**	0.431	0.309	0.298	0.0866
	(0.317)	(0.301)	(0.303)	(0.302)	(0.343)	(0.322)	(0.324)	(0.308)
Evaluator's charact.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluation order dum.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N evaluators	291	291	291	291	291	291	291	291
N Observations	970	970	970	1164	970	970	970	1164

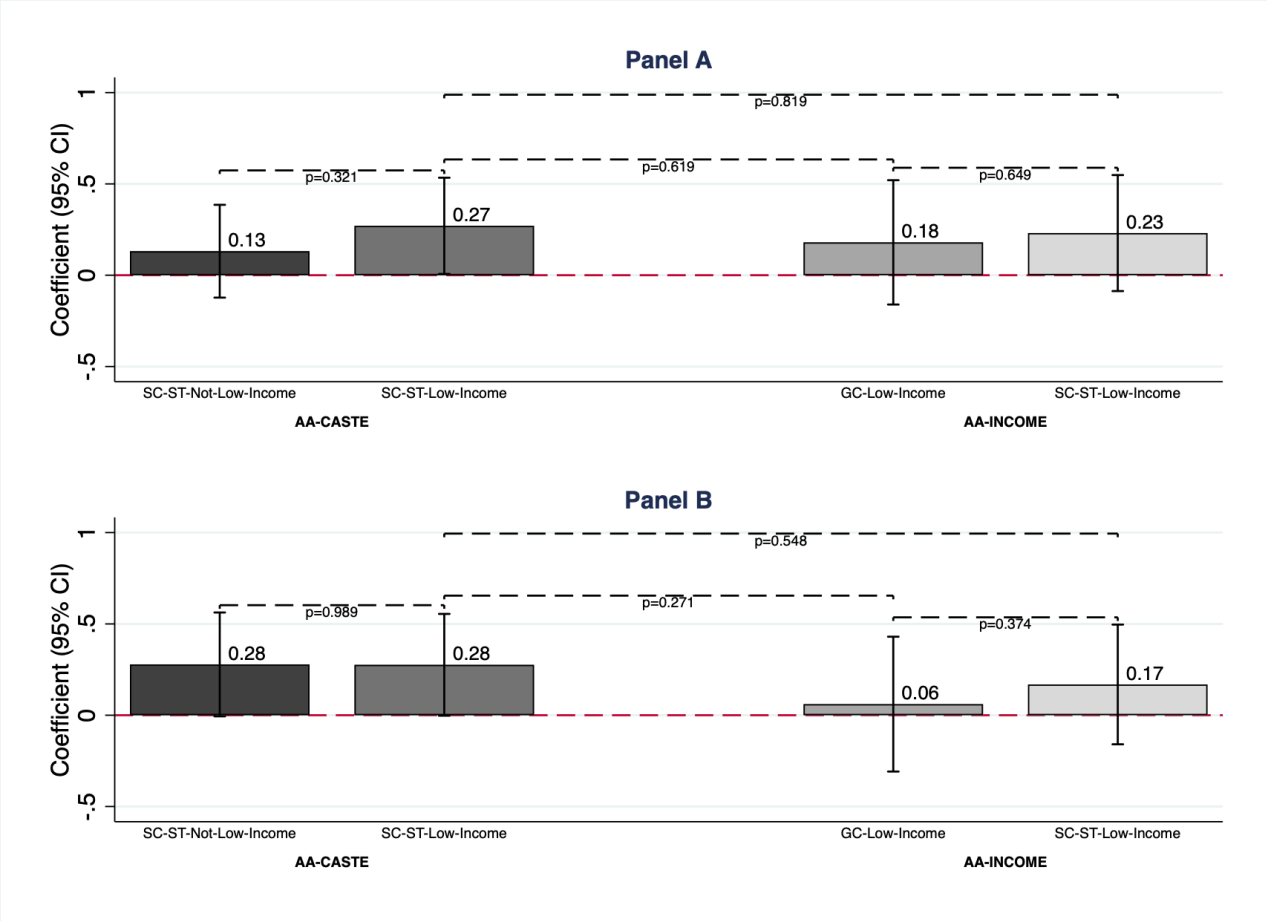
Notes: The dependent variables are the evaluator's standardized share (z-score) of the total bonus allocated to the test-taker in the fourth position in columns (1)–(4) and the standardized difference (z-score) between the guessed contribution of the test taker in the fourth position and the share allocated in columns (5)–(8). Columns (1) and (5) correspond to the baseline specification in Equation (2a), Columns (2) and (6) to the institutional specification in Equation (2b), Columns (3) and (7) to the institutional specification with income interactions in Equation (2c), and Columns (4) and (8) to the extended specification in Equation (3). All regressions include evaluator socio-demographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

to the substantial heterogeneity observed for competence perceptions, estimated allocation effects are generally positive across beneficiary groups, although most estimates are statistically imprecise. Under caste-based affirmative action, both low-income and non-low-income SC-ST beneficiaries receive positive allocation premiums, while under income-based affirmative action both low-income General Category and low-income SC-ST beneficiaries also receive positive premiums. Unlike the competence-perception results in Section 5.2.1, however, pairwise differences across beneficiary groups are not statistically distinguishable.

Allocation relative to perceived competence. Tables 4 and Columns (1)–(4) of Table 5 together suggest a divergence between competence perceptions and redistribution. We therefore next examine deviations between evaluators' guessed competence assessments and their reward allocations. Specifically, Columns (5)–(8) in Table 5 report estimates where the outcome is the difference between the share allocated to a beneficiary and the share implied by the evaluator's own guessed scores. Positive values therefore indicate that evaluators allocate more than their own competence assessments alone would predict.

Under the institutional setting, Column (5) of Table 5 shows that affirmative action ben-

Figure 4: Impact of AA treatment on reward allocation, by income and caste group



Notes. Marginal effects of affirmative action treatments on guessed scores of beneficiaries, by income and caste groups. Marginal effects are computed using the specification in Columns (4) and (8) of Table 5.

eficiaries receive allocations that exceed their implied competence-based shares by 0.236 SD. Column (6) further shows that this compensatory premium is substantially larger for caste-based (0.337 SD units) than for income-based (0.171 SD units affirmative action beneficiaries; in fact the coefficient on AA-Income though positive and economically meaningful is not statistically significant.

Panel B of Figure 4 summarizes the extended-setting results. Under caste-based affirmative action, both low-income and non-low-income SC-ST beneficiaries receive positive deviations of approximately 0.28 SD ($p < 0.10$), indicating that evaluators allocate rewards exceeding what their own competence assessments would predict. Under income-based affirmative action, deviations are considerably smaller and statistically insignificant for both low-income General Category and low-income SC-ST beneficiaries. Taken together, the evidence suggests that caste-based affirmative action generates a compensatory redistributive response that is considerably weaker under income-based affirmative action, although the extended-setting estimates are imprecisely estimated.

5.2.3 Robustness checks

Section A3 of the Appendix presents triple interaction specifications (Tables A3–A5) with all test-takers at all positions included. The findings are robust to this alternative specification. Section A4 of the Appendix presents robustness checks for the three outcomes. Column (1) in Tables A6, A7, and A8 excludes the 13 evaluators who failed at least one attention check. Column (2) excludes the 12 evaluators who allocated the bonus equally across all participants. Columns (3) and (4) split the sample between economics and non-economics students. Restricting the sample as in Columns (1) and (2) produces no substantial changes in the coefficients of interest. By contrast, Columns (3) and (4) highlight differences in behavior towards high-income caste-AA beneficiaries: non-economics students display a stronger stigma of incompetence and allocate less, so that the generosity observed towards this group is primarily driven by economics students (Table A8). Section A5, (Tables A9–A11) reports estimates disaggregated by the caste of the evaluator; the results show that the sample is unfortunately underpowered for this exercise.

6 Discussion and Concluding Comments

Our experimental results reveal a dual pattern in responses to affirmative action in India: beneficiaries from historically marginalized caste groups face stigma in terms of perceived competence, yet this stigma does not translate into corresponding redistribution penalties. Across several specifications, allocations toward caste-based affirmative action beneficiaries are directionally compensatory, although many of these estimates are statistically imprecise. This duality persists even as the framing and rationale for affirmative action shift from caste-based to ostensibly identity-neutral income-based criteria. We discuss the interpretation of these findings, their limitations, and their broader implications.

6.1 Interpretation: What Drives the Dual Response?

Two mechanisms could explain why evaluators simultaneously hold negative perceptions of SC-ST beneficiaries' competence yet do not penalize them in reward allocations.

First, evaluators may recognize that caste and household economic status constitute circumstances of birth beyond individuals' control, and that adverse conditions during formative years constrain access to quality education and resources. Lower academic performance may therefore be interpreted not as reflecting innate ability but as the result of systemic disadvantage. Directionally compensatory allocations toward caste-based affirmative action beneficiaries may reflect a reparative impulse intended to offset disadvantages associated with lower-caste origins.

Second, evaluators may view lower achievement under adverse conditions as evidence of exceptional perseverance and determination. Under this interpretation, an SC-ST student who scores moderately well despite facing discrimination and resource constraints demonstrates greater perseverance than a high-caste student who scores equally well with significant advantages. This would lead evaluators to allocate rewards more generously toward SC-ST beneficiaries even while acknowledging lower measured performance.

Both mechanisms are consistent with our findings and need not be mutually exclusive. What is clear is that evaluators do not simply apply a meritocratic standard based on observed performance. Instead, their allocations appear to incorporate broader considerations related to structural disadvantage and distributive fairness.

Importantly, this compensatory response seem to be stronger under caste-based affirmative action than under income-based affirmative action. General Category beneficiaries of income-based affirmative action are perceived as highly competent and rewarded accordingly, but there is no evidence of additional redistribution beyond what their perceived competence predicts. This suggests that caste remains a more powerful signal of deservingness and structural disadvantage than income alone.

6.2 Why Identity-Neutral Policies Don't Eliminate Bias

A central finding of our study is that even under identical income-based eligibility criteria, SC-ST beneficiaries do not receive the same positive competence appraisal that General Category beneficiaries do. When the policy is framed around economic disadvantage rather than caste, GC beneficiaries receive a substantial competence boost (0.43 SD), while SC-ST beneficiaries receive a much smaller boost (0.22 SD). This 0.21 SD gap demonstrates that economic framing alone cannot overcome deeply embedded social hierarchies and stereotypes about caste and competence.

This has direct policy implications. One argument for replacing caste-based with income-based affirmative action is that it would reduce stigma by framing redistribution around universally recognized disadvantage (poverty) rather than contested social identities. Our evidence suggests this expectation is overly optimistic.

6.3 Limitations and Caveats

Several limitations should be acknowledged. First, our sample consists of university students in three Indian cities. This is an appropriate population for studying attitudes toward affirmative action, as these students have direct experience with reservation policies and are among the key populations directly affected by affirmative action policies. However, generalization to other populations or countries should be made cautiously.

Second, our experiment relies on evaluators making allocation decisions in a controlled setting. While the allocations have real monetary consequences for test-takers, thereby enhancing external validity, evaluators' own payoffs are unaffected by their decisions. Real-world settings such as hiring, promotion, or university admissions often involve stronger competitive pressures and institutional trade-offs that could either amplify or attenuate the patterns we observe.

Third, demand effects are a concern in any experiment on sensitive topics like affirmative action. Evaluators may have inferred the study's purpose and adjusted their responses to appear fair or socially desirable. However, several features of our design mitigate this concern: (1) the online and fully anonymous nature of the experiment limits the scope for socially desirable responding (Mummolo and Peterson; 2019); (2) evaluators were not told the study's focus on affirmative action until debriefing; and (3) identification relies on comparisons across treatments that are symmetrically subject to such pressures. Moreover, the fact that

evaluators openly acknowledged negative stereotypes while simultaneously allocating rewards generously suggests that they were not simply attempting to appear unbiased.

Fourth, our measure of compensatory treatment—the deviation between allocations and guessed scores—assumes that, absent other motives, evaluators would allocate rewards proportionally to perceived competence. This is a simplifying assumption, as evaluators may also consider factors such as equality norms or distributive fairness independently of competence assessments. Accordingly, the deviation measure should be interpreted as capturing departures from competence-based allocations rather than a pure measure of compensatory intent. Nevertheless, the fact that positive deviations are concentrated among caste-based affirmative action beneficiaries, and are considerably weaker under income-based affirmative action, is consistent with evaluators responding differently to the two policy frameworks.

6.4 Conclusion

Affirmative action policies are controversial by their very nature because they improve allocation of preferred positions for certain individuals at the cost of others. Our experimental evidence from India demonstrates that responses to affirmative action are shaped not only by policy design but also by the social identity of beneficiaries. We find that stigma of incompetence is concentrated on beneficiaries from historically marginalized caste groups, and that this stigma persists even when the same individuals qualify through ostensibly identity-neutral income criteria. Yet this stigma does not translate into corresponding material penalties. Instead, allocations toward caste-based affirmative action beneficiaries are directionally compensatory—often exceeding what their perceived competence would predict—whereas such patterns are considerably weaker under income-based affirmative action.

These findings have broader implications for the study of affirmative action and redistribution. The fact that responses to affirmative action are not unidimensional—negative stereotypes about beneficiaries can coexist with compensatory behavior—suggests that stigma and material responses need not move together. This has methodological implications, as measuring beliefs alone (e.g., through surveys or lab experiments) may provide an incomplete or misleading picture of how policies affect real outcomes.

Our findings also carry important implications for policy debates. By showing that different dimensions of disadvantage are not equivalent in shaping perceptions of merit, deservingness, and redistributive justice, our results point to potential limits of class-based alternatives to identity-based policies. Understanding these dynamics is essential for designing policies that effectively address historical disadvantage while minimizing unintended negative consequences.

References

Alesina, A. and Angeletos, G.-M. (2005). Fairness and redistribution, American economic review **95**(4): 960–980.

Almås, I., Cappelen, A. W., Sørensen, E. Ø. and Tungodden, B. (2010). Fairness and the development of inequality acceptance, Science **328**(5982): 1176–1178.

Balafoutas, L., Davis, B. J. and Sutter, M. (2016). Affirmative action or just discrimination? a study on the endogenous emergence of quotas, Journal of Economic Behavior & Organization **127**: 87–98.

Basteck, C. and Mantovani, M. (2018). Cognitive ability and games of school choice, Games and economic behavior **109**: 156–183.

Bobo, L. (1998). Race, interests, and beliefs about affirmative action: Unanswered questions and new directions, American Behavioral Scientist **41**(7): 985–1003.

Bobocel, D. R., Son Hing, L. S., Davey, L. M., Stanley, D. J. and Zanna, M. P. (1998). Justice-based opposition to social policies: Is it genuine?, Journal of Personality and Social Psychology **75**(3): 653–669.

Bors, D. A. and Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form, Educational and Psychological Measurement **58**(3): 382–398.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø. and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach, American Economic Review **97**(3): 818–827.

Cappelen, A. W., Konow, J., Sørensen, E. Ø. and Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness, American Economic Review **103**(4): 1398–1413.

Carpenter, J., Graham, M. and Wolf, J. (2013). Cognitive ability and strategic sophistication, Games and Economic Behavior **80**: 115–130.

Charness, G., Rustichini, A. and Van de Ven, J. (2018). Self-confidence and strategic behavior, Experimental Economics **21**: 72–98.

Chen, D. L., Schonger, M. and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments, Journal of Behavioral and Experimental Finance **9**: 88–97.

Danz, D., Vesterlund, L. and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility, American Economic Review **112**(9): 2851–2883.

Deshpande, A. (2019). Double jeopardy? stigma of identity and affirmative action, The Review of Black Political Economy **46**(1): 38–64.

Deshpande, A. and Ramachandran, R. (2022). Viewing caste inequality upside down: The perversity of special schemes for brahmins in south india, Economic and Political Weekly **57**(32): 43–49.

Dorrough, A. R., Leszczyńska, M., Barreto, M. and Glöckner, A. (2016). Revealing side effects of quota rules on group cooperation, Journal of Economic Psychology **57**: 136–152.

Federico, C. M. and Sidanius, J. (2002). Racism, ideology, and affirmative action revisited: The antecedents and consequences of "principled objections" to affirmative action, Journal of Personality and Social Psychology **82**(4): 488–502.

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution, Journal of Public Economics **82**(2): 225–246.

Frederick, S. (2005). Cognitive reflection and decision making, Journal of Economic Perspectives pp. 25–42.

Gangadharan, L., Jain, T., Maitra, P. and Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders, European Economic Review **90**: 302–325.

Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis, Journal of Political Economy **124**(6): 1619–1676.

Gille, V. (2022). Experimental Evidence on Affirmative Action, Springer Nature Singapore, Singapore, pp. 1–10.
URL: *https://doi.org/10.1007/978-981-33-4016-9˙37-1*

Heilman, M. E., Block, C. J. and Lucas, J. A. (1992). Presumed incompetent? stigmatization and affirmative action efforts., Journal of applied psychology **77**(4): 536.

Heilman, M. E., Block, C. J. and Stathatos, P. (1997). The affirmative action stigma of incompetence: Effects of performance information ambiguity, Academy of Management Journal **40**(3): 603–625.

Herzog, S., Schildberg-Hörisch, H., Trieu, C. and Willrodt, J. (2023). Who is in favor of affirmative action? representative evidence from an experiment and a survey, IZA Discussion Paper No. 16640 .

Hossain, T. and Okui, R. (2013). The binarized scoring rule, Review of Economic Studies **80**(3): 984–1001.

Kinder, D. R. and Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life, Journal of Personality and Social Psychology **40**(3): 414–431.

Kluegel, J. R. and Smith, E. R. (1986). Beliefs About Inequality: Americans' Views of What Is and What Ought to Be, Aldine de Gruyter, New York.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions, American economic review **90**(4): 1072–1092.

Leibbrandt, A., Wang, L. C. and Foo, C. (2018). Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women, Management Science **64**(8): 3501–3516.

Mummolo, J. and Peterson, E. (2019). Demand effects in survey experiments: An empirical assessment, American Political Science Review **113**(2): 517–529.

Petters, L. M. and Schroeder, M. (2020). Negative side effects of affirmative action: How quotas lead to distortions in performance evaluation, European Economic Review **130**: 103500.

Proto, E., Rustichini, A. and Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions, Journal of Political Economy **127**(3): 1351–1390.

Proto, E., Rustichini, A. and Sofianos, A. (2022). Intelligence, errors, and cooperation in repeated interactions, The Review of Economic Studies **89**(5): 2723–2767.

Raven, J. C. (1962). Advanced Progressive Matrices: Set II, London: H. K. Lewis.

Reyna, C., Henry, P. J., Korfmacher, W. and Tucker, A. (2006). Examining the principles in principled conservatism: The role of responsibility stereotypes as cues for deservingness in racial policy decisions, Journal of Personality and Social Psychology **90**(1): 109–128.

Schildberg-Hörisch, H., Schwarz, M. A., Trieu, C. and Willrodt, J. (2023). Perceive fairness and consequences of affirmative action policies, The Economic Journal **133**: 3099–3135.

Sears, D. O. and Henry, P. J. (2003). The origins of symbolic racism, Journal of Personality and Social Psychology **85**(2): 259–275.

Sniderman, P. M. and Carmines, E. G. (1997). Reaching Beyond Race, Harvard University Press, Cambridge, MA.

Son Hing, L. S., Bobocel, D. R. and Zanna, M. P. (2002). Meritocracy and opposition to affirmative action: Making concessions in the face of discrimination, Journal of Personality and Social Psychology **83**(3): 493–509.

Tarman, C. and Sears, D. O. (2005). The conceptualization and measurement of symbolic racism, Journal of Politics **67**(3): 731–761.

Thorat, A. and Joshi, O. (2020). The continuing practice of untouchability in india, Economic & Political Weekly **55**(2): 37.

Online Appendix

Index of Appendix Tables and Figures

Appendix Tables

Appendix Figures

A1 Sample of test-takers

Table A1: Characteristics of test-takers

	(1) Mean	(2) Std. dev.
Age	21.72	(2.80)
Female	0.56	(0.50)
SC	0.09	(0.29)
OBC	0.21	(0.41)
General category	0.70	(0.46)
Hindu	0.82	(0.38)
Muslim	0.07	(0.26)
Other religion	0.11	(0.31)
Low-income HH	0.25	(0.43)
Nb people HH	4.23	(1.58)
Governement school	0.30	(0.46)
Private School	0.70	(0.46)
Year educ	16.68	(2.33)
Econ field	0.58	(0.50)
N Test-takers	57	

Notes: This table reports the mean values and standard deviations for a set of characteristics for our sample of 57 test-takers.

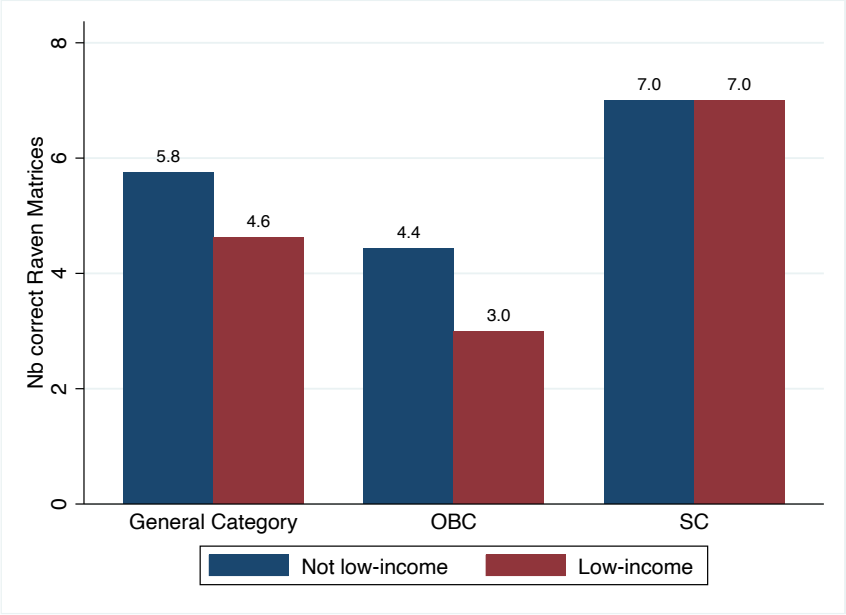


Figure A1: Number of correct Raven matrices by caste and income status

Notes. This figure reports the number of Raven matrices, out of 12, solved correctly on average by our sample of test-takers, by caste and income category.

A2 Regression table for control group

Table A2: Identity effects on guessed scores and allocations – Control group only.

	Guessed Score (1)	Allocation (2)	Deviation (3)
SC-ST	-0.359***	0.142*	0.331***
	(0.0851)	(0.0733)	(0.0840)
Low-Income	-0.248***	0.312***	0.458***
	(0.0675)	(0.0746)	(0.0936)
Constant	0.0441	0.725***	-0.00741
	(0.202)	(0.0968)	(0.101)
Evaluator's characteristics	Yes	No	No
Grade dummies	Yes	Yes	Yes
Position dummies	Yes	Yes	Yes
Round dummy	Yes	No	No
Evaluation order dummies	Yes	Yes	Yes
N evaluators	102	102	102
N Observations	1632	1632	1632

Notes: This tables reports the estimates of equation (1). The dependent variable in column (1) is the evaluator's guessed score for the test-taker in the fourth position. Raw guessed scores are rescaled to a common 0–25 scale across grade levels and then standardized (z-scores). The dependent variable in column (2) is the evaluator's standardized share (z-score) of the total bonus allocated to the test-taker in the fourth position. The dependent variable in column (3) is the standardized difference (z-score) between the guessed contribution of the test taker in the fourth position and the share allocated. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A3 Robustness checks: triple interaction with all participants

Table A3: Impact of AA treatment on guessed score –triple interaction

	Outcome: Guessed Score			
	Institutional			Extended
	(1)	(2)	(3)	(4)
SC-ST	-0.266***	-0.232***	-0.266***	-0.232***
	(0.0451)	(0.0430)	(0.0478)	(0.0433)
Low-Income	-0.119***	-0.111***	-0.135***	-0.136***
	(0.0359)	(0.0353)	(0.0387)	(0.0367)
4^{th} position × Low-Income	0.0415	-0.0299	-0.207	-0.189
	(0.0812)	(0.0813)	(0.141)	(0.140)
Low-Income × SC-ST			0.104	0.0987
			(0.0777)	(0.0782)
AA treatment	0.112**			
	(0.0537)			
AA-Caste		0.113*	0.108*	0.111*
		(0.0611)	(0.0617)	(0.0621)
AA-Income		0.104	0.0972	0.0548
		(0.0669)	(0.0668)	(0.0626)
AA treatment × 4^{th} position	-0.133*			
	(0.0799)			
AA-Caste × 4^{th} position		-0.186**	-0.296**	-0.291**
		(0.0883)	(0.118)	(0.118)
AA-Caste × 4^{th} position × Low-Income			0.183	0.177
			(0.133)	(0.133)
AA-Income × 4^{th} position		0.00973	0.112	0.175
		(0.0976)	(0.123)	(0.124)
AA-Income × 4^{th} position × SC-ST				-0.00494
				(0.124)
Constant	0.142	0.135	0.142	0.0972
	(0.182)	(0.184)	(0.185)	(0.184)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Grade dummies	Yes	Yes	Yes	Yes
Position dummies	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	291	291	291	291
N Observations	4900	4900	4900	5820

Notes: The dependent variable is the evaluator's guessed score. Raw guessed scores are rescaled to a common 0–25 scale across grade levels and then standardized (z-scores). All regressions include evaluator socio-demographic characteristics.Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Impact of AA treatment on reward allocation –triple interaction

	Outcome: Share Allocation			
	Institutional			Extended
	(1)	(2)	(3)	(4)
SC-ST	0.0500	0.0500	0.0489	0.0635
	(0.0513)	(0.0395)	(0.0440)	(0.0426)
Low-Income	0.278***	0.278***	0.270***	0.270***
	(0.0433)	(0.0432)	(0.0402)	(0.0443)
4^{th} position × Low-Income	-0.0368	-0.0368	-0.152	-0.113
	(0.0613)	(0.0915)	(0.138)	(0.142)
Low-Income × SC-ST			0.00583	-0.0297
			(0.0785)	(0.0746)
AA treatment	-0.103***			
	(0.0311)			
AA-Caste		-0.103***	-0.102***	-0.101***
		(0.0328)	(0.0346)	(0.0344)
AA-Income		-0.103**	-0.101**	-0.0728*
		(0.0455)	(0.0452)	(0.0424)
AA treatment × 4^{th} position	0.274**			
	(0.124)			
AA-Caste × 4^{th} position		0.274**	0.155	0.147
		(0.130)	(0.115)	(0.114)
AA-Caste × 4^{th} position × Low-Income			0.178	0.192
			(0.120)	(0.119)
AA-Income × 4^{th} position		0.274	0.318	0.271
		(0.201)	(0.218)	(0.207)
AA-Income × 4^{th} position × SC-ST				0.00938
				(0.129)
Constant	1.170***	1.170***	1.168***	1.152***
	(0.0539)	(0.0537)	(0.0551)	(0.0525)
Grade dummies	Yes	Yes	Yes	Yes
Position dummies	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	291	291	291	291
N Observations	3920	3920	3920	4656

Notes: The dependent variable is the evaluator's standardized share (z-score) of the total bonus allocated to test-takers. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Impact of AA treatment on reward allocation (deviation from guessed score) –triple interaction

	Outcome: Share All. - Share Guessed Score			
	Institutional			Extended
	(1)	(2)	(3)	(4)
SC-ST	0.201***	0.201***	0.230***	0.236***
	(0.0592)	(0.0496)	(0.0584)	(0.0531)
Low-Income	0.435***	0.437***	0.447***	0.449***
	(0.0614)	(0.0614)	(0.0609)	(0.0651)
4^{th} position × Low-Income	-0.122	-0.121	-0.0241	-0.00595
	(0.0943)	(0.109)	(0.180)	(0.181)
Low-Income × SC-ST			-0.0940	-0.112
			(0.110)	(0.104)
AA treatment	-0.109***			
	(0.0375)			
AA-Caste		-0.117***	-0.113***	-0.112***
		(0.0390)	(0.0413)	(0.0410)
AA-Income		-0.0876	-0.0860	-0.0666
		(0.0568)	(0.0566)	(0.0515)
AA treatment × 4^{th} position	0.377**			
	(0.151)			
AA-Caste × 4^{th} position		0.388**	0.391**	0.388**
		(0.155)	(0.173)	(0.172)
AA-Caste × 4^{th} position × Low-Income			-0.00928	-0.00400
			(0.165)	(0.164)
AA-Income × 4^{th} position		0.354	0.285	0.253
		(0.245)	(0.259)	(0.244)
AA-Income × 4^{th} position × SC-ST				-0.0155
				(0.156)
Constant	-0.0312	-0.0338	-0.0402	-0.0513
	(0.0617)	(0.0615)	(0.0637)	(0.0611)
Grade dummies	Yes	Yes	Yes	Yes
Position dummies	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	291	291	291	291
N Observations	3920	3920	3920	4656

Notes: The dependent variable is the standardized difference (z-score) between the guessed contribution of the test-takers and the shares allocated. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A4 Robustness checks: sample restriction and heterogeneous effects

Table A6: Robustness checks for guessed scores

	Outcome: Guessed Score			
	Attention	Alloc. uneq	Non-econ	Econ
	(1)	(2)	(3)	(4)
Low-Income	-0.355**	-0.318**	-0.359	-0.336**
	(0.139)	(0.140)	(0.291)	(0.146)
AA-Caste	-0.306**	-0.222	-0.360	-0.252*
	(0.136)	(0.136)	(0.263)	(0.147)
AA-Caste × Low-Income	0.342**	0.279*	0.645**	0.161
	(0.148)	(0.149)	(0.317)	(0.151)
AA-Income	0.429***	0.464***	0.764***	0.313**
	(0.117)	(0.116)	(0.220)	(0.131)
AA-Income × SC-ST	-0.193*	-0.211*	-0.475*	-0.0946
	(0.116)	(0.118)	(0.263)	(0.119)
Constant	0.106	0.109	-0.493	0.427
	(0.322)	(0.306)	(0.983)	(0.326)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	278	279	87	204
N Observations	1112	1116	348	816

Notes: The dependent variable is the evaluator's guessed score. Raw guessed scores are rescaled to a common 0–25 scale across grade levels and then standardized (z-scores). All columns correspond to the extended specification in Equation (2c). Column (1) excludes the evaluators who failed at least one attention check. Column (2) excludes the evaluators who allocated the bonus equally across all participants. Columns (3) and (4) split the sample between economics and non-economics students. All regressions include evaluator sociodemographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Robustness checks for rewards allocations

	Outcome: Allocation			
	Attention	Alloc. uneq	Non-econ	Econ
	(1)	(2)	(3)	(4)
Low-Income	0.149	0.205	0.117	0.218
	(0.127)	(0.128)	(0.191)	(0.160)
AA-Caste	0.0910	0.145	-0.125	0.220
	(0.134)	(0.134)	(0.203)	(0.168)
AA-Caste × Low-Income	0.195	0.142	0.314	0.0744
	(0.144)	(0.145)	(0.253)	(0.171)
AA-Income	0.185	0.198	0.294	0.116
	(0.180)	(0.179)	(0.229)	(0.227)
AA-Income × SC-ST	0.0287	0.0474	-0.163	0.140
	(0.115)	(0.115)	(0.290)	(0.101)
Constant	-0.582*	-0.628**	-0.489	-0.636*
	(0.313)	(0.310)	(1.017)	(0.346)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	278	279	87	204
N Observations	1112	1116	348	816

Notes: The dependent variable is the evaluator's standardized share (z-score) of the total bonus allocated to the four selected test-takers. All columns correspond to the extended specification in Equation (2c). Column (1) excludes the evaluators who failed at least one attention check. Column (2) excludes the evaluators who allocated the bonus equally across all participants. Columns (3) and (4) split the sample between economics and non-economics students. All regressions include evaluator socio-demographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Robustness checks for rewards allocations (Deviation from guessed score)

	Outcome: Share All. - Share Guessed Score			
	Attention	Alloc. uneq	Non-econ	Econ
	(1)	(2)	(3)	(4)
Low-Income	0.292*	0.338**	0.246	0.369**
	(0.150)	(0.151)	(0.247)	(0.182)
AA-Caste	0.244	0.274*	0.0670	0.351*
	(0.150)	(0.150)	(0.237)	(0.185)
AA-Caste × Low-Income	0.0484	0.00903	0.0768	-0.0355
	(0.170)	(0.171)	(0.315)	(0.196)
AA-Income	0.0622	0.0716	0.117	0.00991
	(0.195)	(0.194)	(0.254)	(0.245)
AA-Income × SC-ST	0.0820	0.108	0.0263	0.142
	(0.124)	(0.125)	(0.292)	(0.121)
Constant	0.141	0.102	0.713	-0.0507
	(0.323)	(0.320)	(1.021)	(0.357)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	278	279	87	204
N Observations	1112	1116	348	816

Notes: The dependent variable is the difference between the guessed contribution of the test taker and the evaluator's allocation for the four selected test-takers. All columns correspond to the extended specification in Equation (2c). Column (1) excludes the evaluators who failed at least one attention check. Column (2) excludes the evaluators who allocated the bonus equally across all participants. Columns (3) and (4) split the sample between economics and non-economics students. All regressions include evaluator socio-demographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A5 Robustness checks: in group-bias?

Table A9: Robustness checks for guessed scores

	Outcome: Guessed Score			
	SC-ST	OBC	SC-ST-OBC	GC
	(1)	(2)	(3)	(4)
Low-Income	-0.811	-0.0591	-0.304	-0.362**
	(0.506)	(0.293)	(0.252)	(0.160)
AA-Caste	-0.184	-0.575	-0.385	-0.216
	(0.413)	(0.351)	(0.268)	(0.154)
AA-Caste × Low-Income	0.198	0.142	0.153	0.362**
	(0.518)	(0.261)	(0.240)	(0.178)
AA-Income	0.662	0.374	0.433*	0.448***
	(0.396)	(0.315)	(0.232)	(0.130)
AA-Income × SC-ST	-0.00268	-0.452	-0.293	-0.168
	(0.232)	(0.368)	(0.258)	(0.129)
Constant	2.980*	-0.320	0.299	-0.149
	(1.687)	(1.002)	(0.821)	(0.325)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	28	57	85	206
N Observations	112	228	340	824

Notes: The dependent variable is the evaluator's guessed score for participants in the fourth position. All columns correspond to the extended specification in Equation (2c). Column (1) excludes the evaluators who failed at least one attention check. Column (2) excludes the evaluators who allocated the bonus equally across all participants. Columns (3) and (4) split the sample between economics and non-economics students. All regressions include evaluator socio-demographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Robustness checks for rewards allocations

	Outcome: Allocation			
	SC-ST	OBC	SC-ST-OBC	GC
	(1)	(2)	(3)	(4)
Low-Income	0.115	0.120	0.118	0.237
	(0.432)	(0.258)	(0.221)	(0.149)
AA-Caste	0.190	-0.252	-0.0372	0.157
	(0.381)	(0.335)	(0.245)	(0.153)
AA-Caste × Low-Income	0.510	0.102	0.236	0.0920
	(0.363)	(0.269)	(0.213)	(0.179)
AA-Income	0.390	0.293	0.342	0.0926
	(0.339)	(0.329)	(0.248)	(0.216)
AA-Income × SC-ST	-0.218	0.225	0.0803	0.0427
	(0.409)	(0.226)	(0.200)	(0.135)
Constant	-0.981	-0.785	-0.825	-0.376
	(1.328)	(0.795)	(0.641)	(0.309)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	28	57	85	206
N Observations	112	228	340	824

Notes: The dependent variable is the evaluator's allocation for the four selected test-takers. All columns correspond to the extended specification in Equation (2c). Column (1) excludes the evaluators who failed at least one attention check. Column (2) excludes the evaluators who allocated the bonus equally across all participants. Columns (3) and (4) split the sample between economics and non-economics students. All regressions include evaluator socio-demographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Robustness checks for rewards allocations (Deviation from grade)

	Outcome: Share All. - Share Guessed Score			
	SC-ST	OBC	SC-ST-OBC	GC
	(1)	(2)	(3)	(4)
Low-Income	0.343	0.149	0.212	0.400**
	(0.542)	(0.298)	(0.264)	(0.173)
AA-Caste	0.296	0.0368	0.176	0.274
	(0.432)	(0.331)	(0.252)	(0.176)
AA-Caste × Low-Income	0.474	-0.00342	0.156	-0.0781
	(0.513)	(0.312)	(0.265)	(0.205)
AA-Income	0.230	0.208	0.244	-0.0416
	(0.397)	(0.367)	(0.275)	(0.234)
AA-Income × SC-ST	-0.198	0.406	0.207	0.0698
	(0.424)	(0.287)	(0.234)	(0.143)
Constant	-0.0737	0.0314	0.00336	0.358
	(1.263)	(0.830)	(0.664)	(0.329)
Evaluator's characteristics	Yes	Yes	Yes	Yes
Round dummy	Yes	Yes	Yes	Yes
Evaluation order dummies	Yes	Yes	Yes	Yes
N evaluators	28	57	85	206
N Observations	112	228	340	824

Notes: The dependent variable is the difference between the guessed contribution of the test taker and the evaluator's allocation for the four selected test-takers. All columns correspond to the extended specification in Equation (2c). Column (1), and A8 excludes the evaluators who failed at least one attention check. Column (2) excludes the evaluators who allocated the bonus equally across all participants. Columns (3) and (4) split the sample between economics and non-economics students. All regressions include evaluator socio-demographic characteristics, round dummies and dummies for the evaluation order of the group. Robust standard errors are clustered at the evaluator level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A6 Experimental Instructions

A6.1 Test-taker instructions

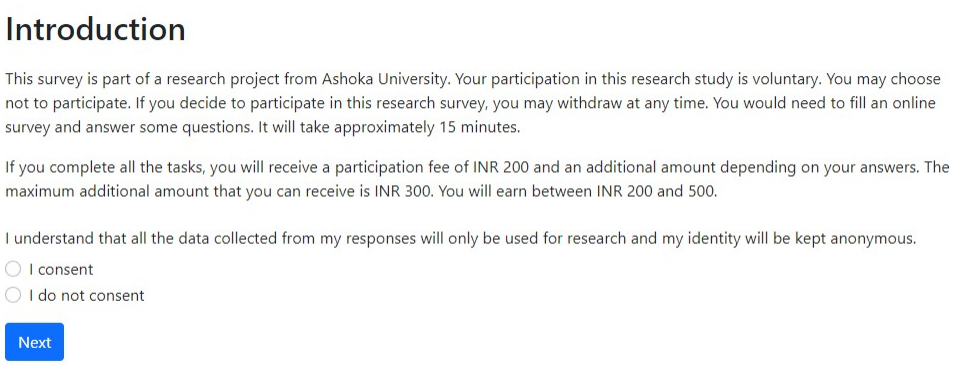


Figure A2: Introduction & Consent

Notes. Experimental screen shown to test-takers.

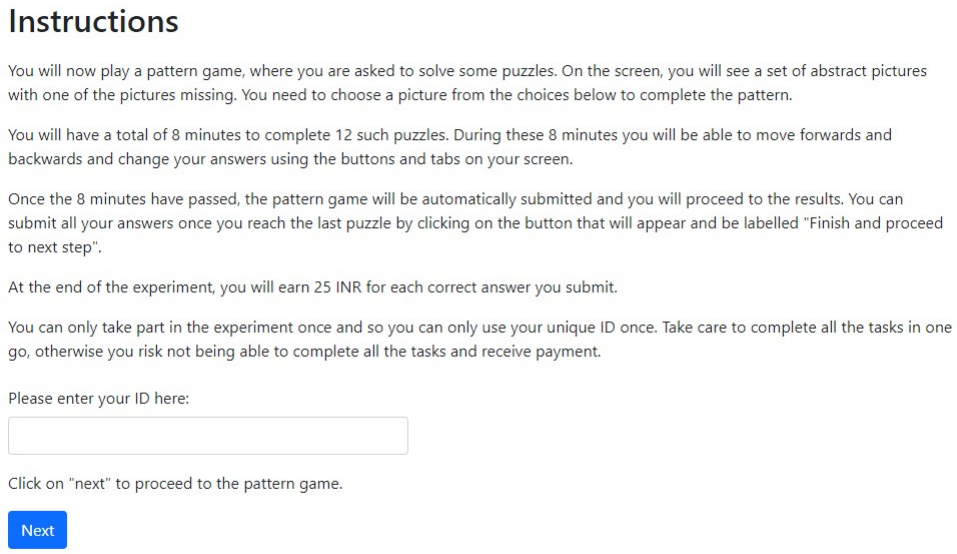


Figure A3: Raven Instructions

Notes. Experimental screen shown to test-takers.

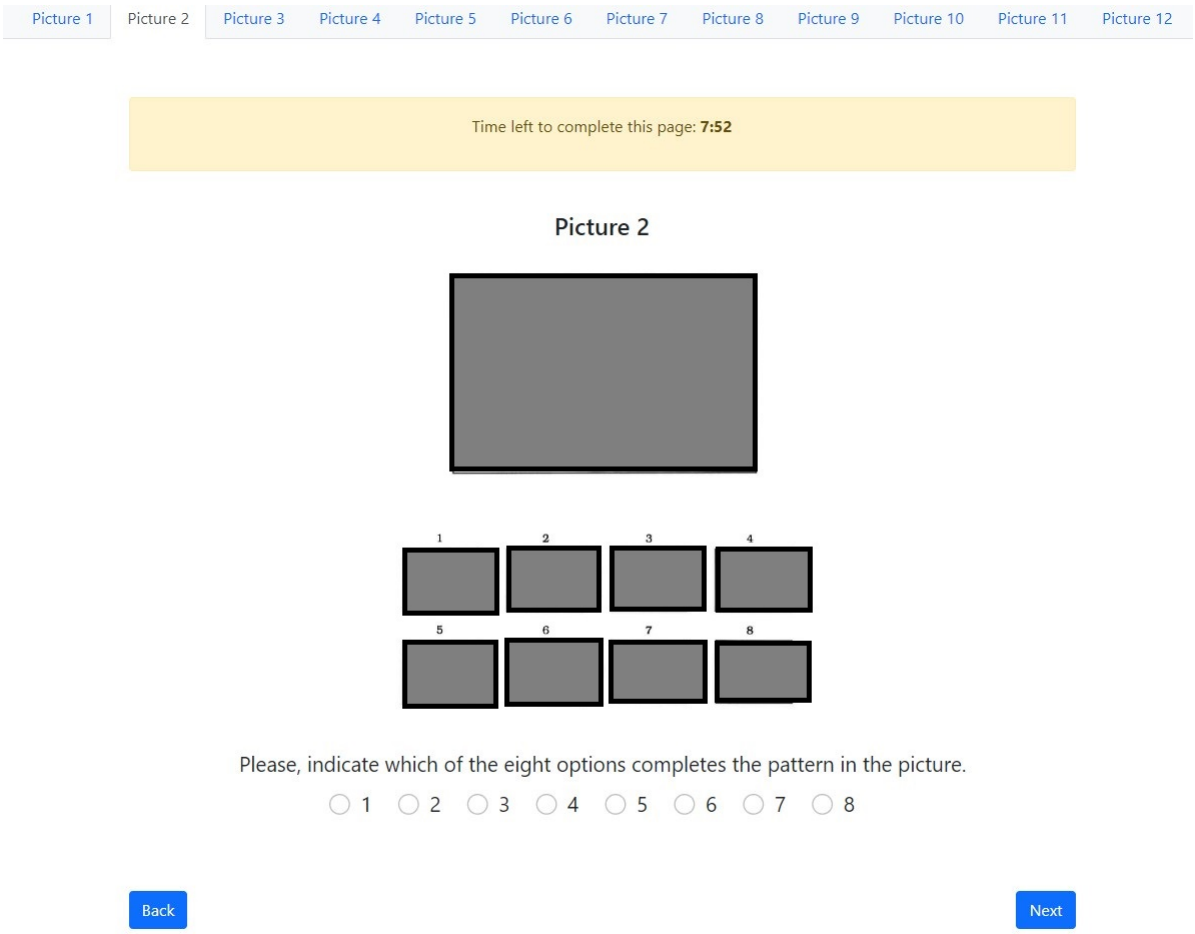


Figure A4: Raven Screen (the Raven matrix is not printed as the images are copyrighted)

Notes. Experimental screen shown to test-takers.

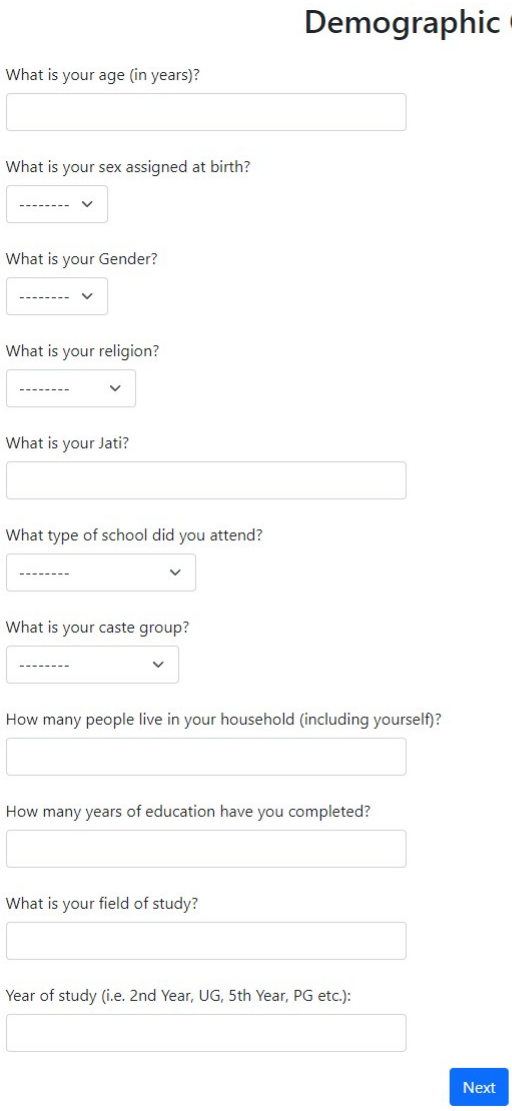


Figure A5: Demographics Questionnaire - Part 1

Notes. Experimental screen shown to test-takers.

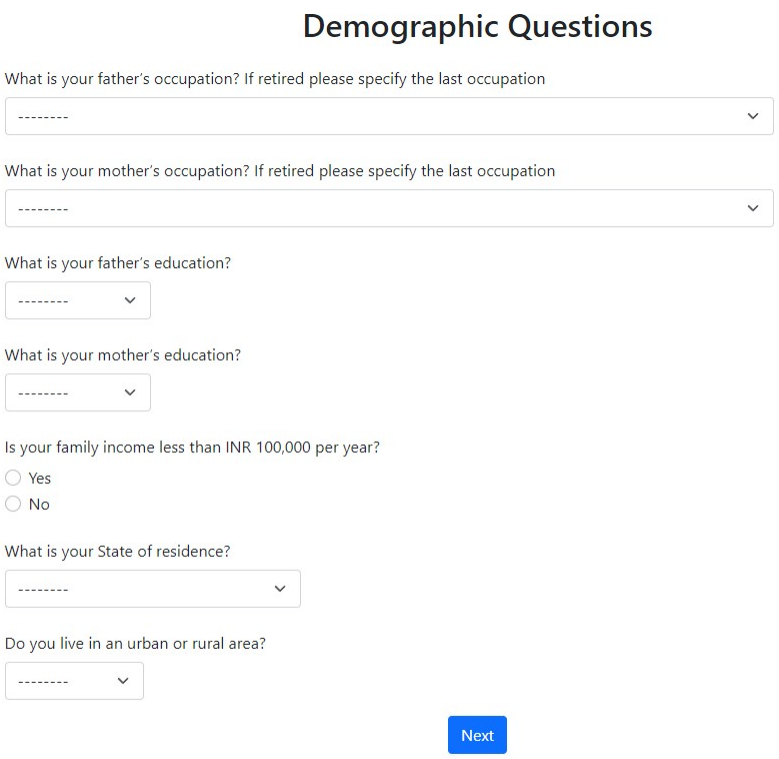


Figure A6: Demographics Questionnaire - Part 2

Notes. Experimental screen shown to test-takers.

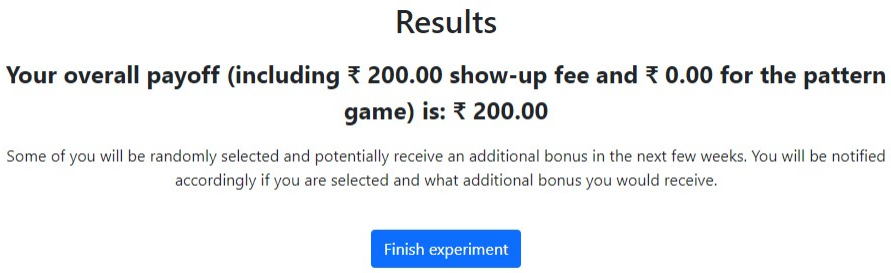


Figure A7: appendix˙exhb Page

Notes. Experimental screen shown to test-takers.

A6.2 Evaluator instructions

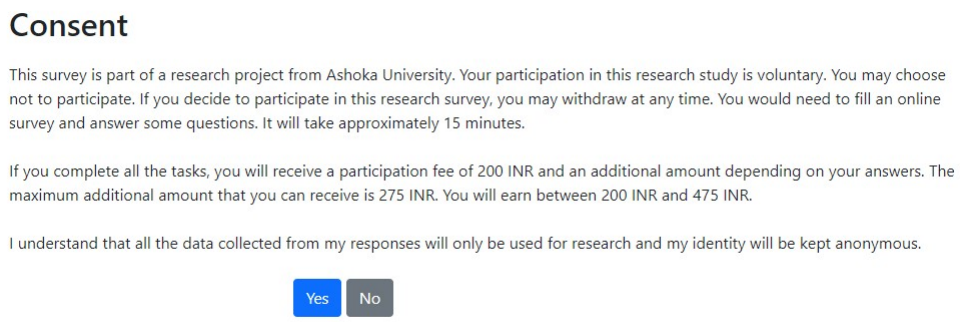


Figure A8: Introduction & Consent

Notes. Experimental screen shown to evaluators.

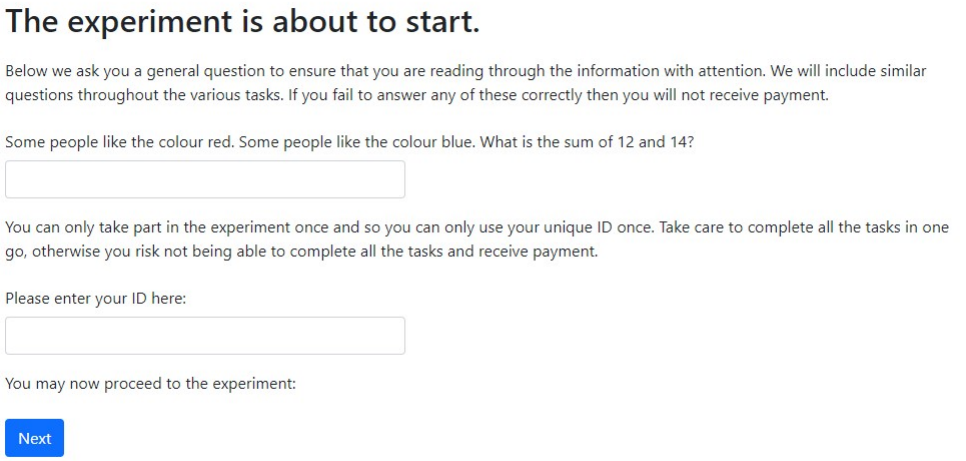


Figure A9: Introduction & Attention Check

Notes. Experimental screen shown to evaluators.

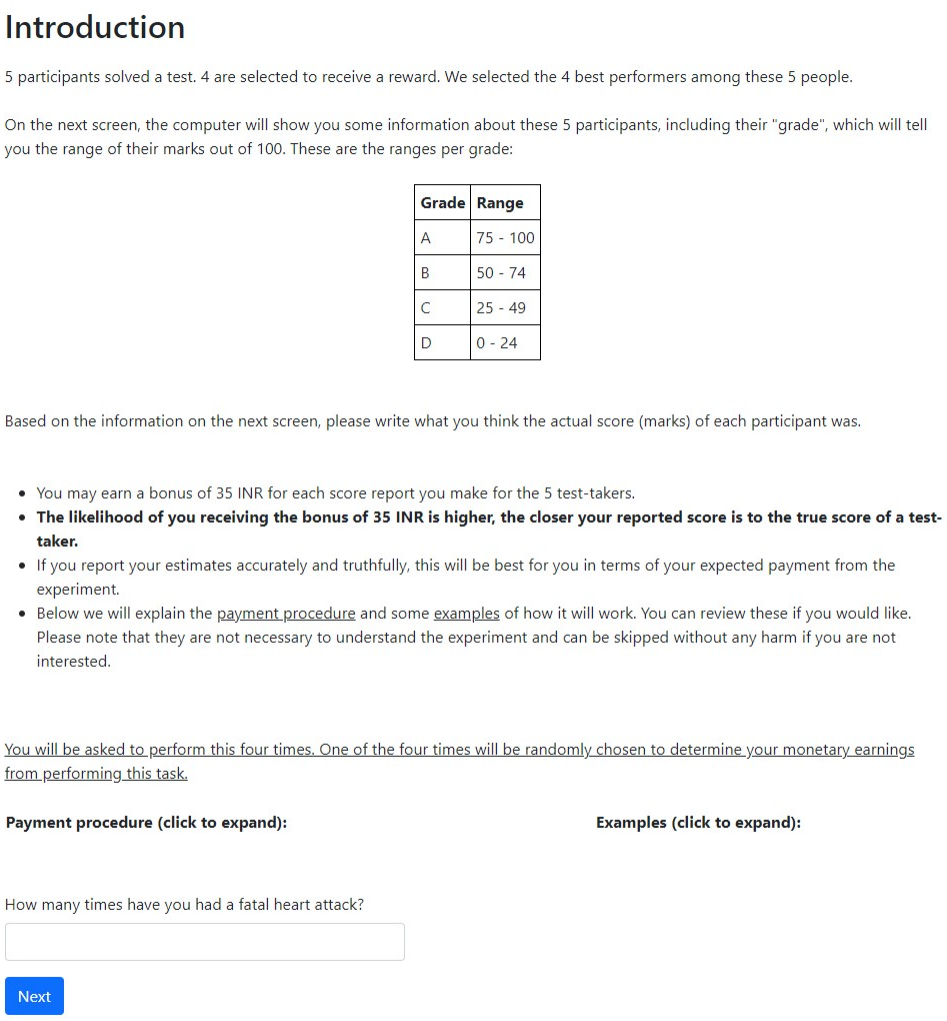
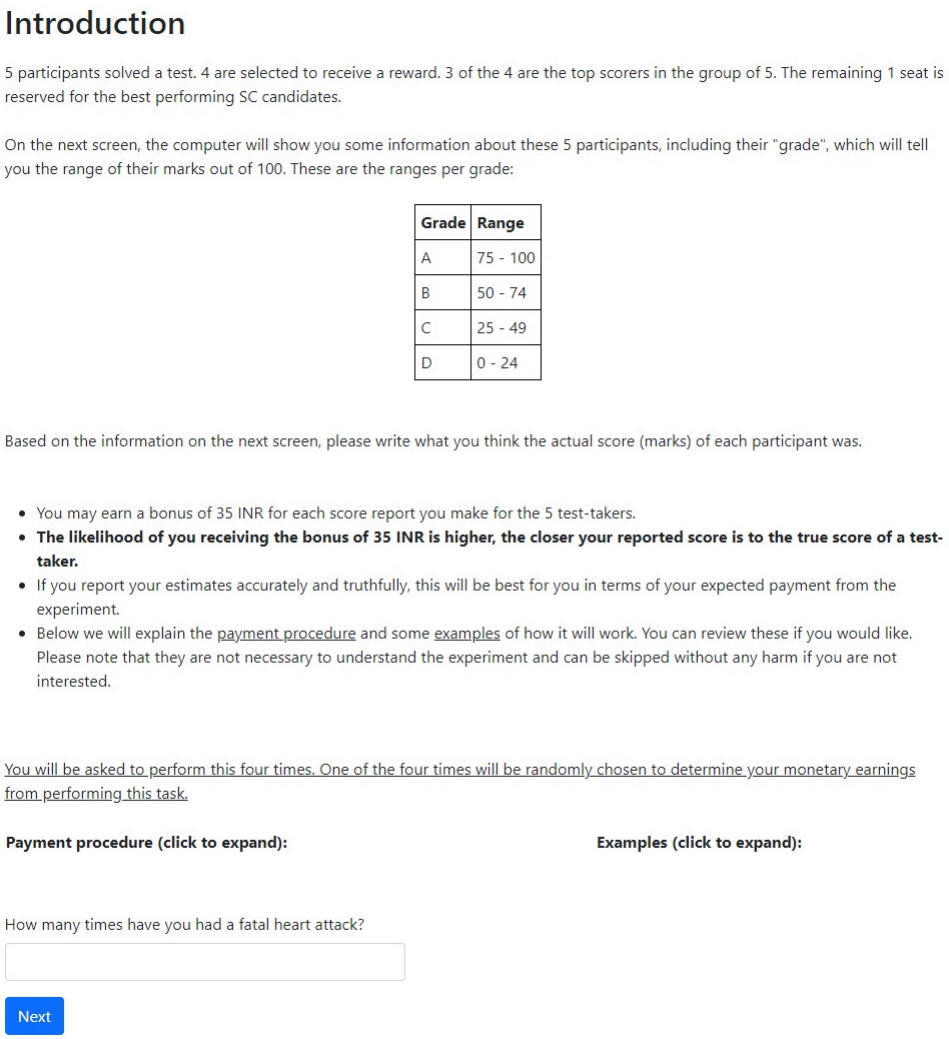


Figure A10: Score guessing introduction - Control

Notes. Experimental screen shown to evaluators.



Introduction

5 participants solved a test. 4 are selected to receive a reward. 3 of the 4 are the top scorers in the group of 5. The remaining 1 seat is reserved for the best performing SC candidates.

On the next screen, the computer will show you some information about these 5 participants, including their "grade", which will tell you the range of their marks out of 100. These are the ranges per grade:

Grade	Range
A	75 - 100
B	50 - 74
C	25 - 49
D	0 - 24

Based on the information on the next screen, please write what you think the actual score (marks) of each participant was.

- You may earn a bonus of 35 INR for each score report you make for the 5 test-takers.
- **The likelihood of you receiving the bonus of 35 INR is higher, the closer your reported score is to the true score of a test-taker.**
- If you report your estimates accurately and truthfully, this will be best for you in terms of your expected payment from the experiment.
- Below we will explain the payment procedure and some examples of how it will work. You can review these if you would like. Please note that they are not necessary to understand the experiment and can be skipped without any harm if you are not interested.

You will be asked to perform this four times. One of the four times will be randomly chosen to determine your monetary earnings from performing this task.

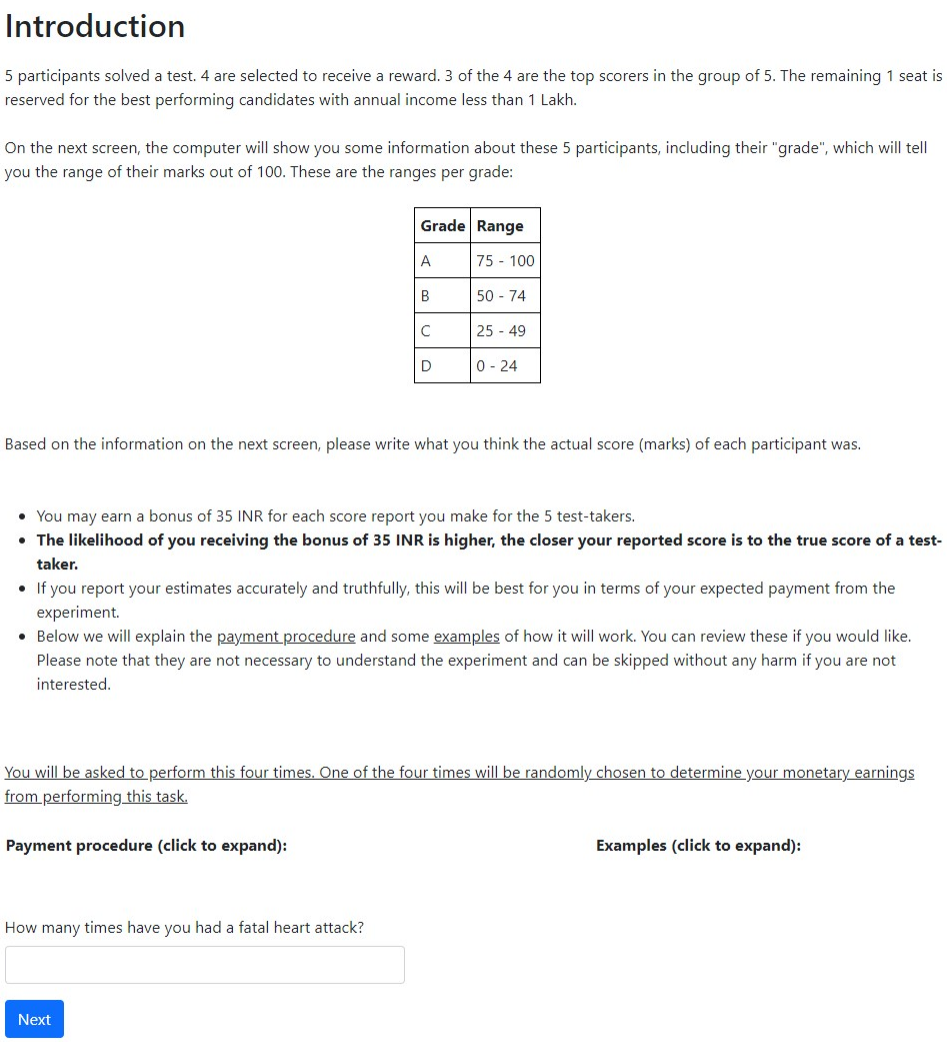
Payment procedure (click to expand): **Examples (click to expand):**

How many times have you had a fatal heart attack?

Next

Figure A11: Score guessing introduction - AA-Caste

Notes. Experimental screen shown to evaluators.



Introduction

5 participants solved a test. 4 are selected to receive a reward. 3 of the 4 are the top scorers in the group of 5. The remaining 1 seat is reserved for the best performing candidates with annual income less than 1 Lakh.

On the next screen, the computer will show you some information about these 5 participants, including their "grade", which will tell you the range of their marks out of 100. These are the ranges per grade:

Grade	Range
A	75 - 100
B	50 - 74
C	25 - 49
D	0 - 24

Based on the information on the next screen, please write what you think the actual score (marks) of each participant was.

- You may earn a bonus of 35 INR for each score report you make for the 5 test-takers.
- **The likelihood of you receiving the bonus of 35 INR is higher, the closer your reported score is to the true score of a test-taker.**
- If you report your estimates accurately and truthfully, this will be best for you in terms of your expected payment from the experiment.
- Below we will explain the payment procedure and some examples of how it will work. You can review these if you would like. Please note that they are not necessary to understand the experiment and can be skipped without any harm if you are not interested.

You will be asked to perform this four times. One of the four times will be randomly chosen to determine your monetary earnings from performing this task.

Payment procedure (click to expand): **Examples (click to expand):**

How many times have you had a fatal heart attack?

Next

Figure A12: Score guessing introduction - AA-Income

Notes. Experimental screen shown to evaluators.

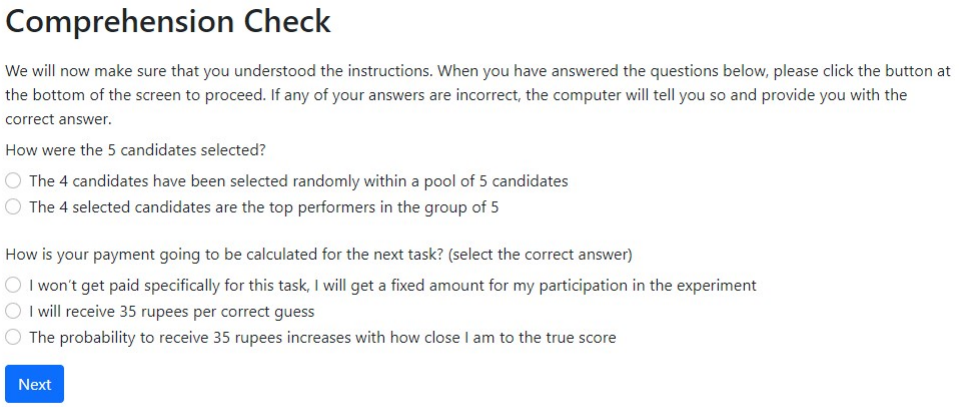


Figure A13: Score guessing quiz - Control

Notes. Experimental screen shown to evaluators.

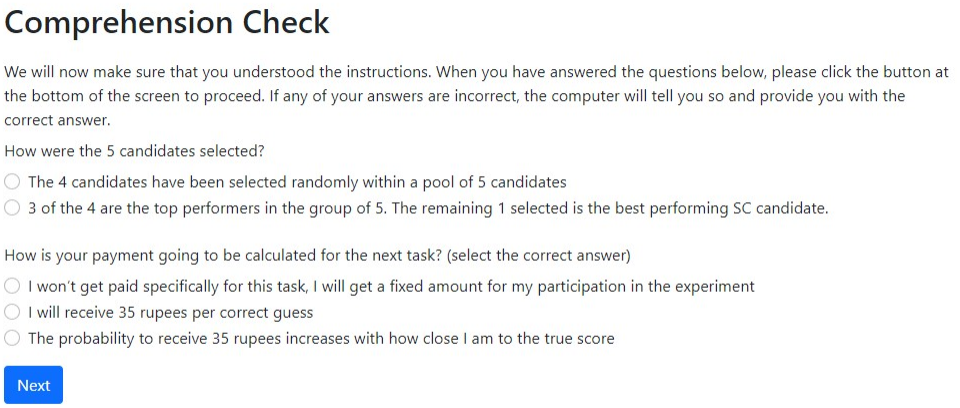


Figure A14: Score guessing quiz - AA-Caste

Notes. Experimental screen shown to evaluators.

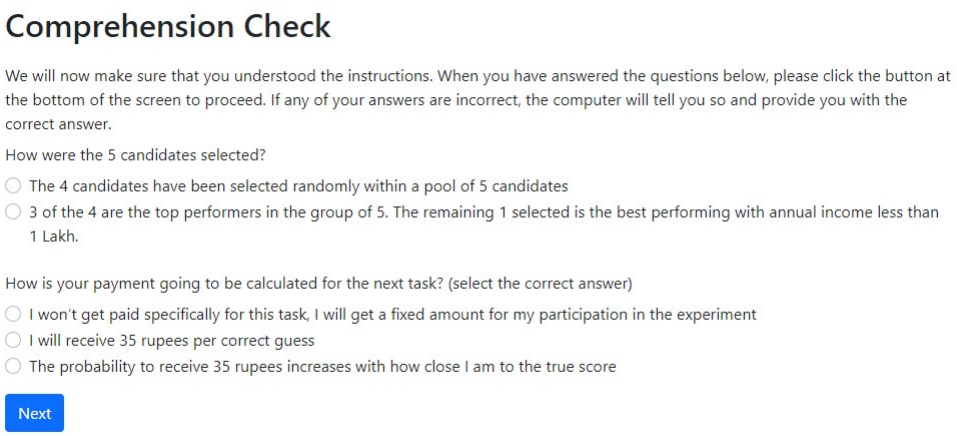


Figure A15: Score guessing quiz - AA-Income

Notes. Experimental screen shown to evaluators.

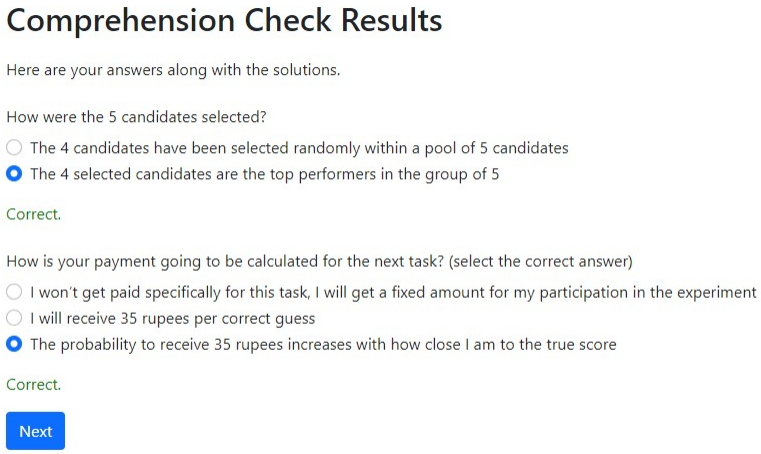


Figure A16: Score guessing quiz feedback - Control

Notes. Experimental screen shown to evaluators.

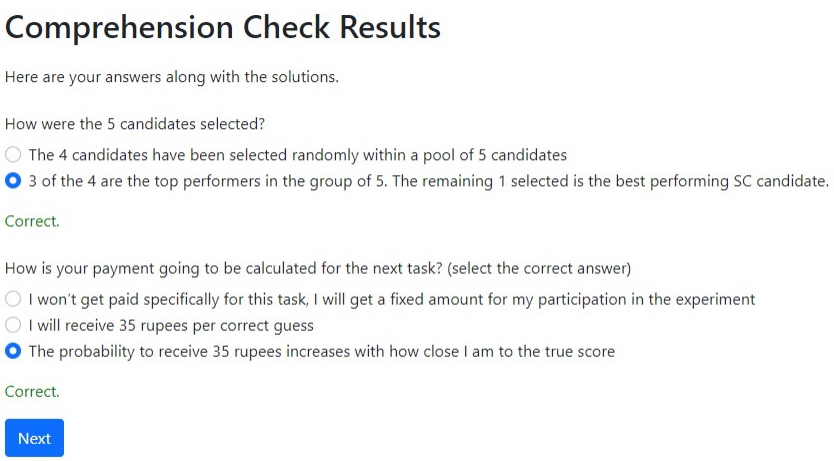


Figure A17: Score guessing quiz feedback - AA-Caste

Notes. Experimental screen shown to evaluators.

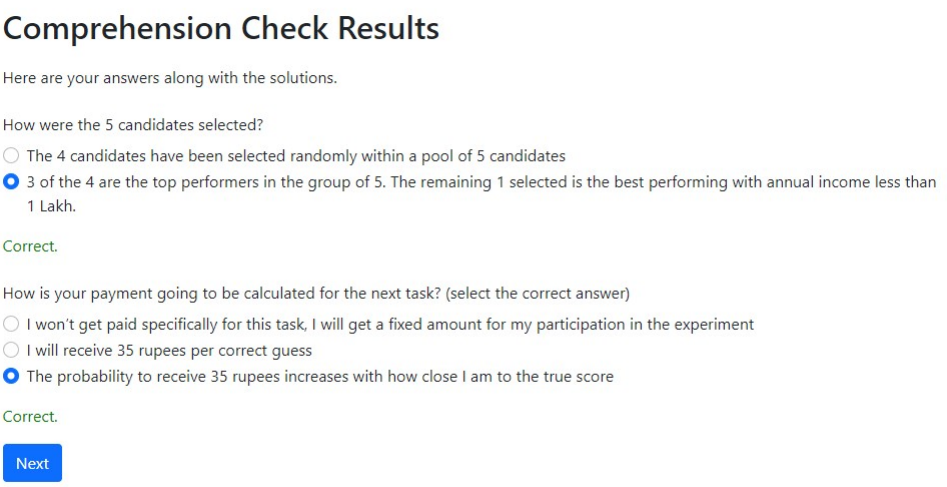


Figure A18: Score guessing quiz feedback - AA-Income

Notes. Experimental screen shown to evaluators.

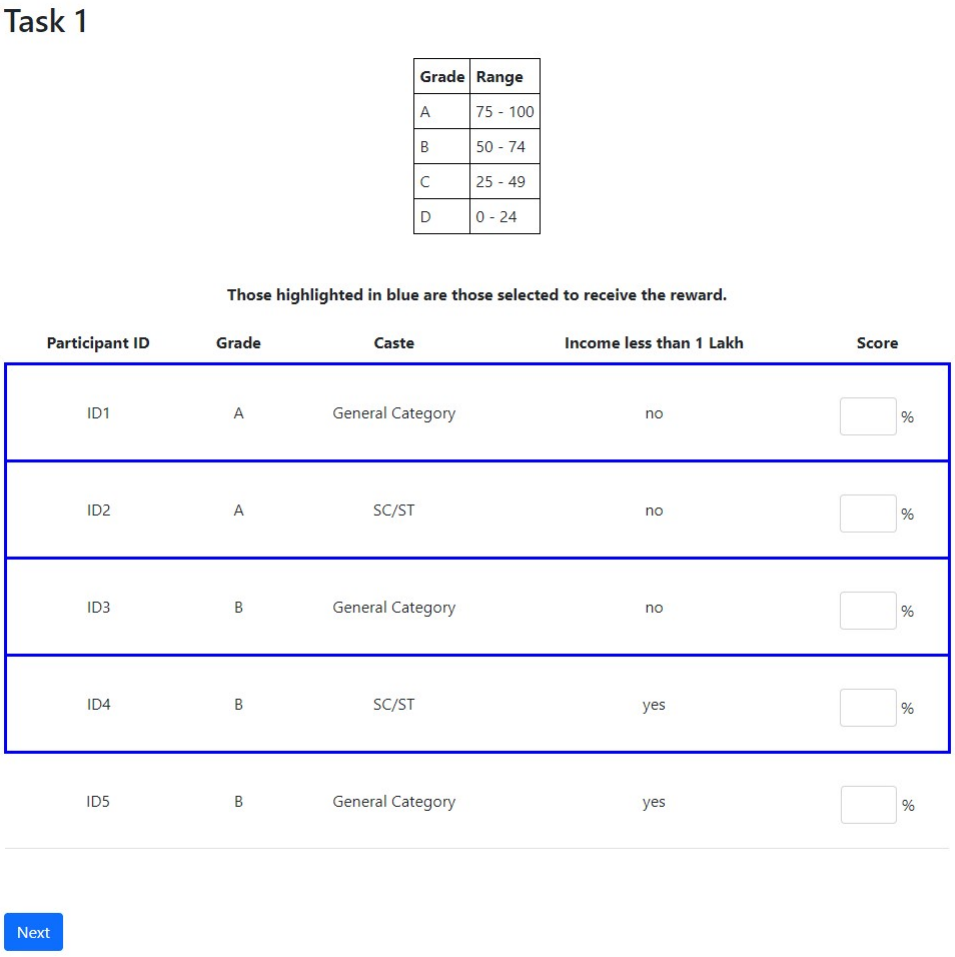
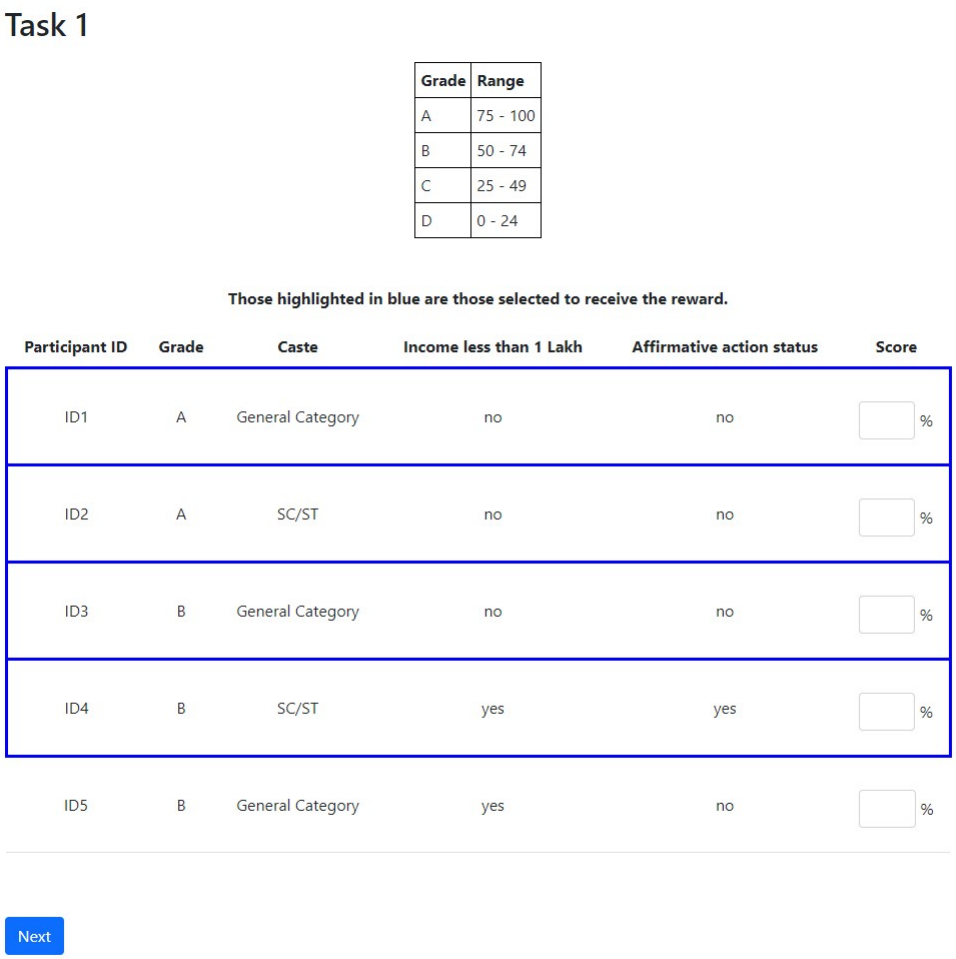


Figure A19: Score guessing: Group 1 - Control

Notes. Experimental screen shown to evaluators.



Task 1

Grade	Range
A	75 - 100
B	50 - 74
C	25 - 49
D	0 - 24

Those highlighted in blue are those selected to receive the reward.

Participant ID	Grade	Caste	Income less than 1 Lakh	Affirmative action status	Score
ID1	A	General Category	no	no	%
ID2	A	SC/ST	no	no	%
ID3	B	General Category	no	no	%
ID4	B	SC/ST	yes	yes	%
ID5	B	General Category	yes	no	%

Next

Figure A20: Score guessing: Group 1 - AA-Caste

Notes. Experimental screen shown to evaluators.

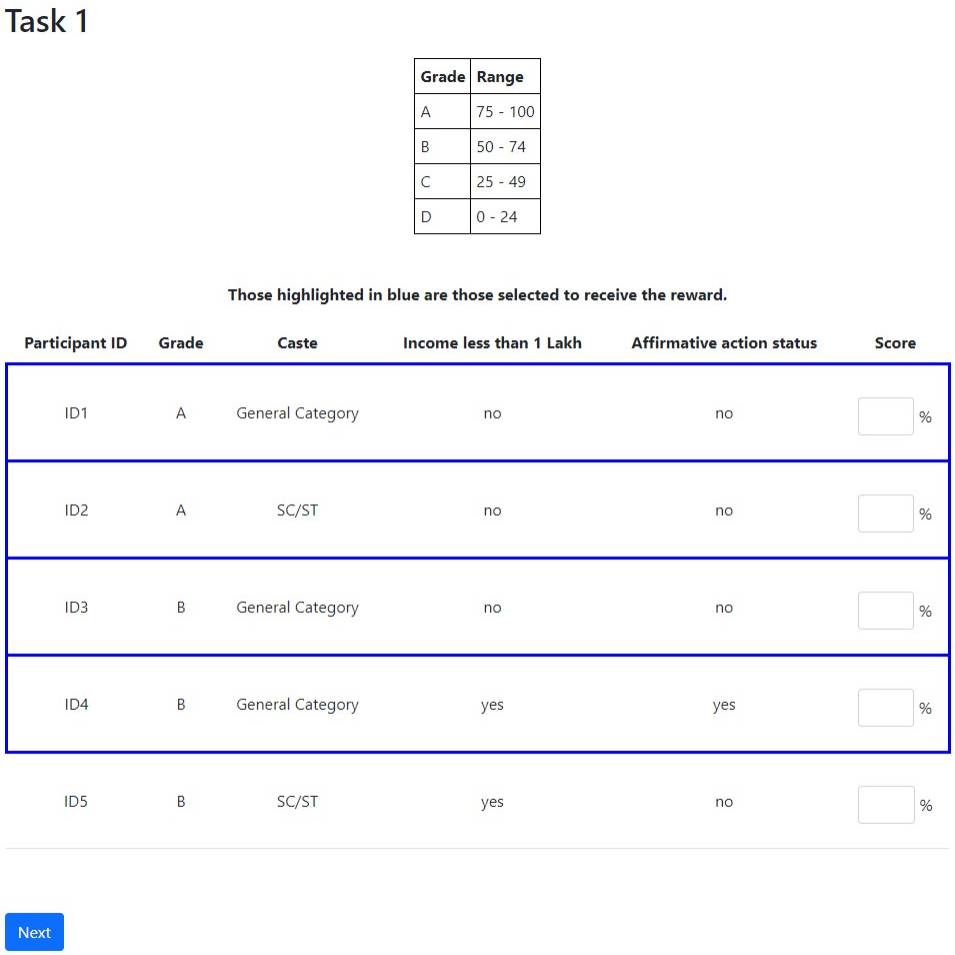


Figure A21: Score guessing: Group 1 - AA-Income

Notes. Experimental screen shown to evaluators.

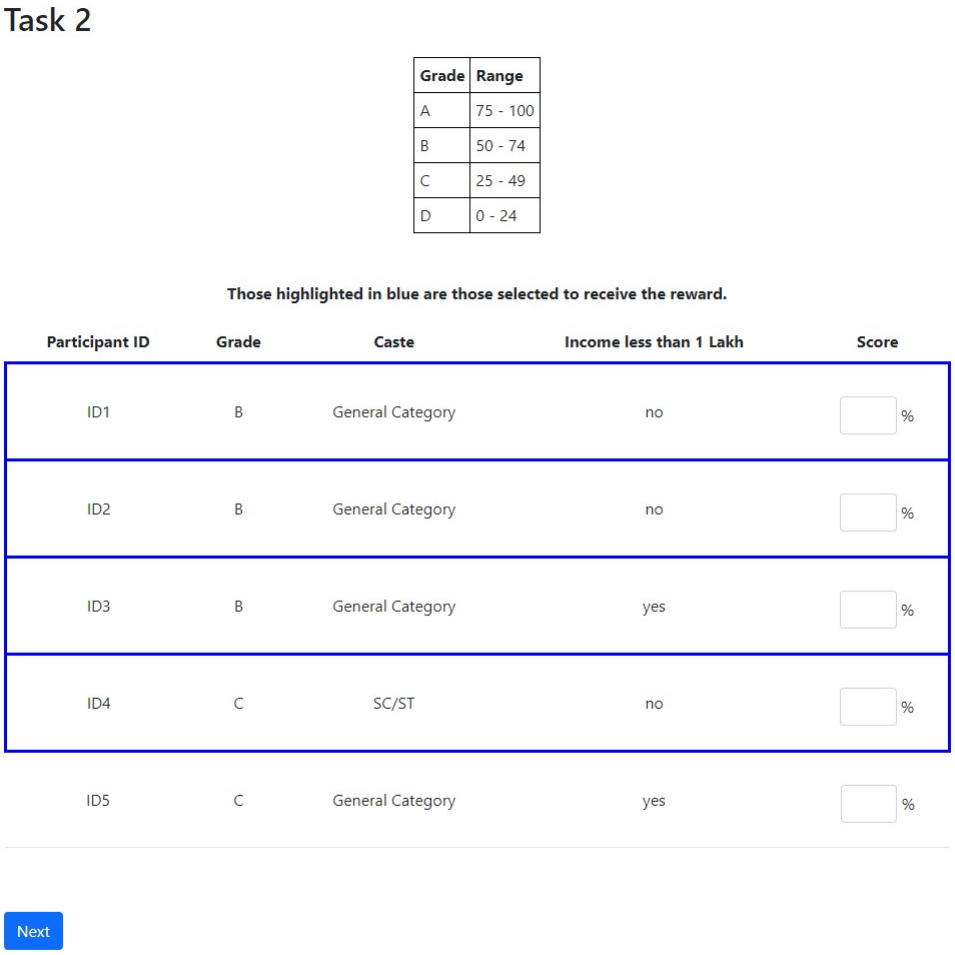
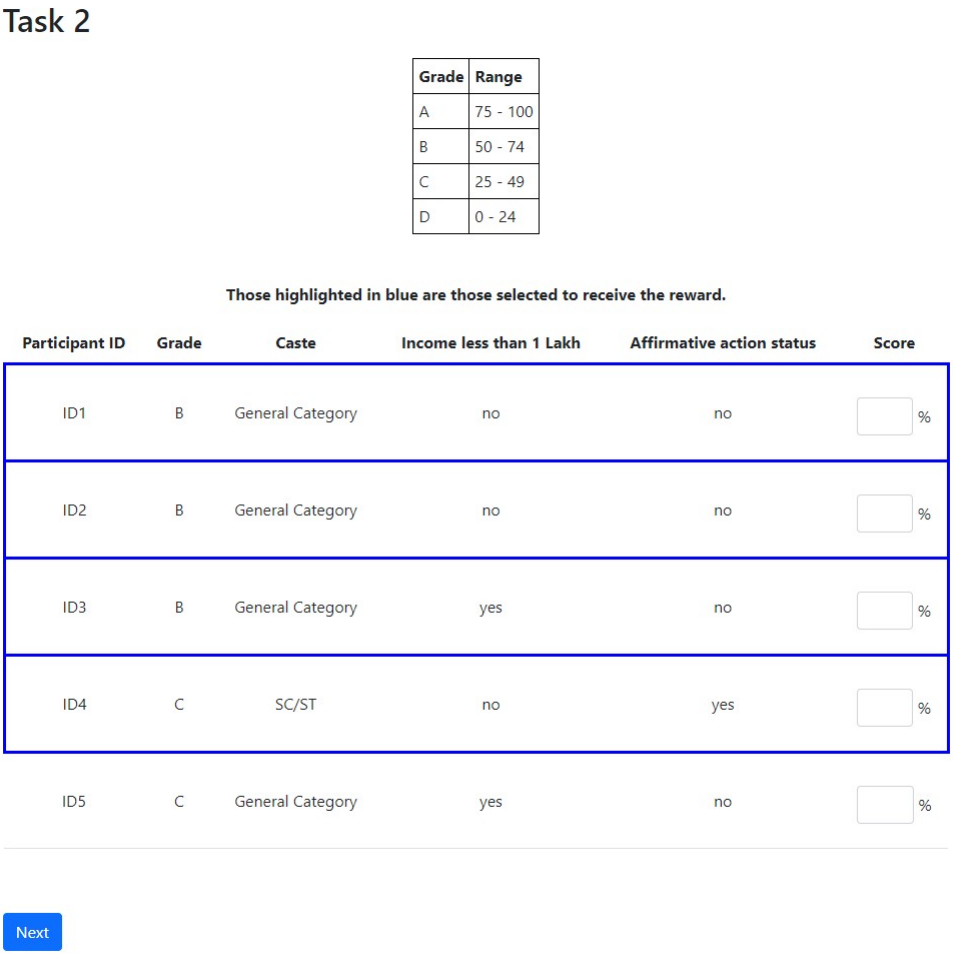


Figure A22: Score guessing: Group 2 - Control

Notes. Experimental screen shown to evaluators.



Task 2

Grade	Range
A	75 - 100
B	50 - 74
C	25 - 49
D	0 - 24

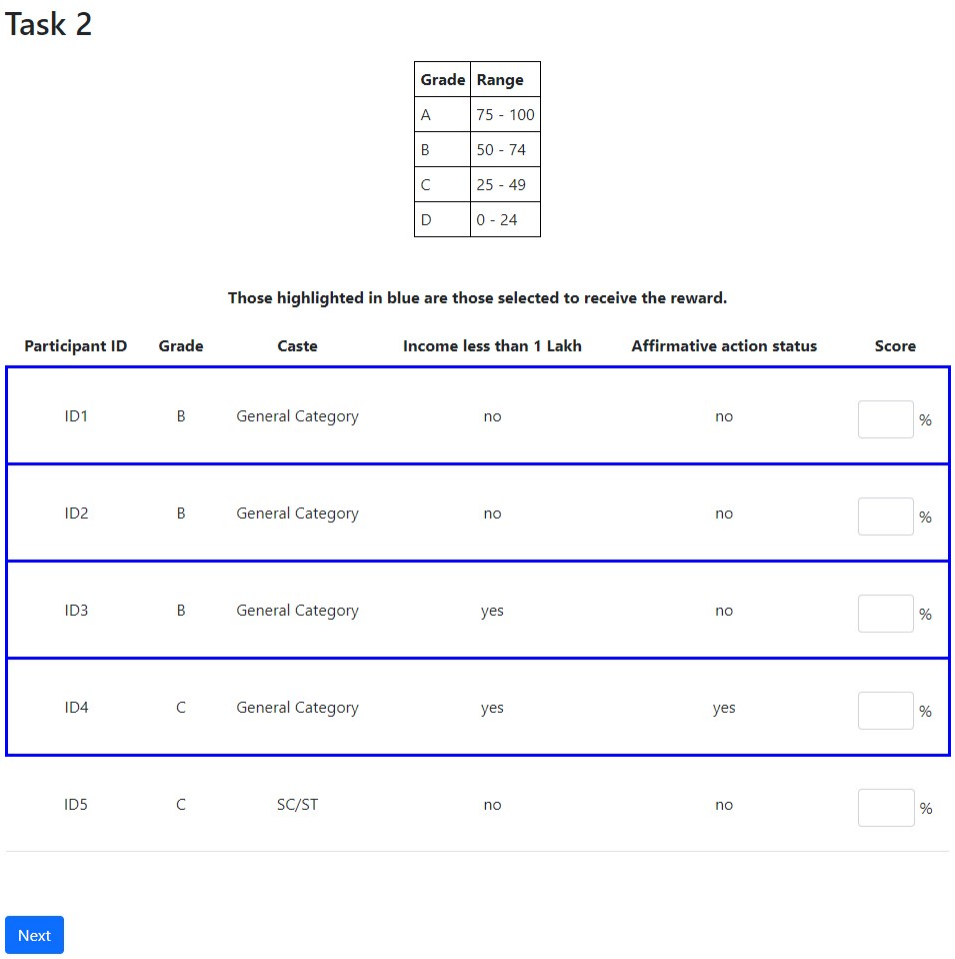
Those highlighted in blue are those selected to receive the reward.

Participant ID	Grade	Caste	Income less than 1 Lakh	Affirmative action status	Score
ID1	B	General Category	no	no	%
ID2	B	General Category	no	no	%
ID3	B	General Category	yes	no	%
ID4	C	SC/ST	no	yes	%
ID5	C	General Category	yes	no	%

Next

Figure A23: Score guessing: Group 2 - AA-Caste

Notes. Experimental screen shown to evaluators.



Task 2

Grade	Range
A	75 - 100
B	50 - 74
C	25 - 49
D	0 - 24

Those highlighted in blue are those selected to receive the reward.

Participant ID	Grade	Caste	Income less than 1 Lakh	Affirmative action status	Score
ID1	B	General Category	no	no	%
ID2	B	General Category	no	no	%
ID3	B	General Category	yes	no	%
ID4	C	General Category	yes	yes	%
ID5	C	SC/ST	no	no	%

Next

Figure A24: Score guessing: Group 2 - AA-Income

Notes. Experimental screen shown to evaluators.

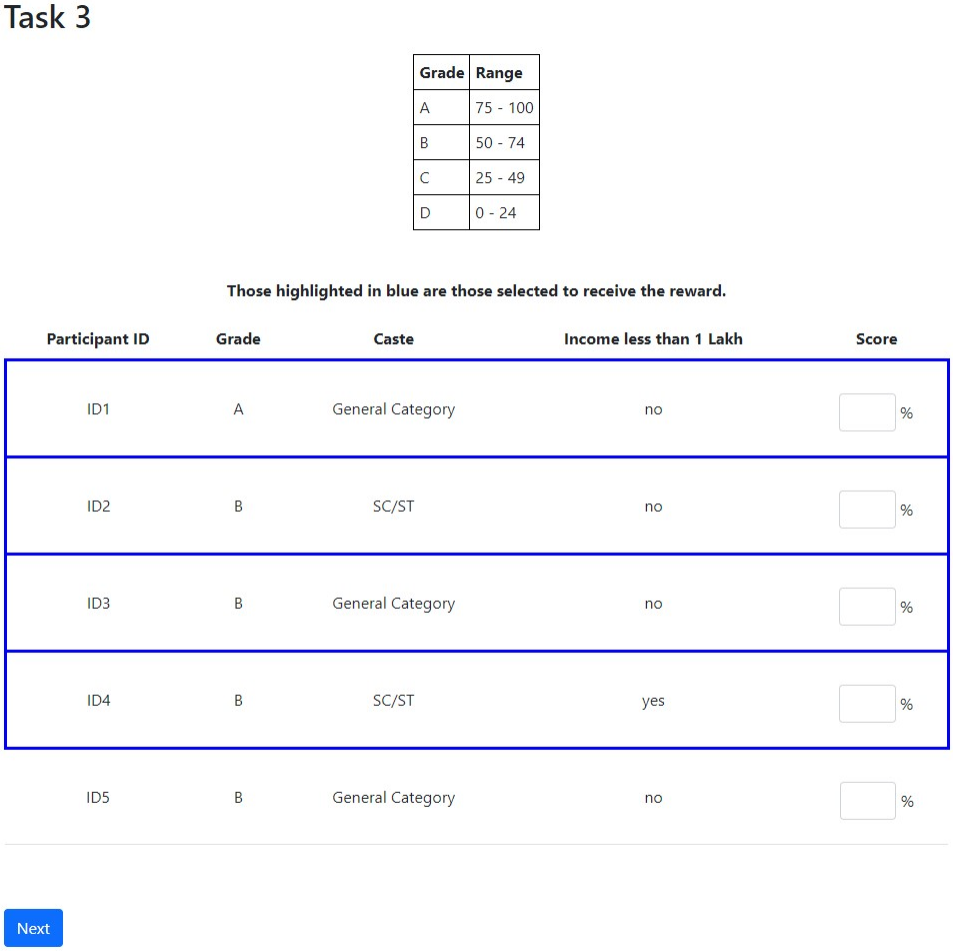


Figure A25: Score guessing: Group 3 - Control

Notes. Experimental screen shown to evaluators.

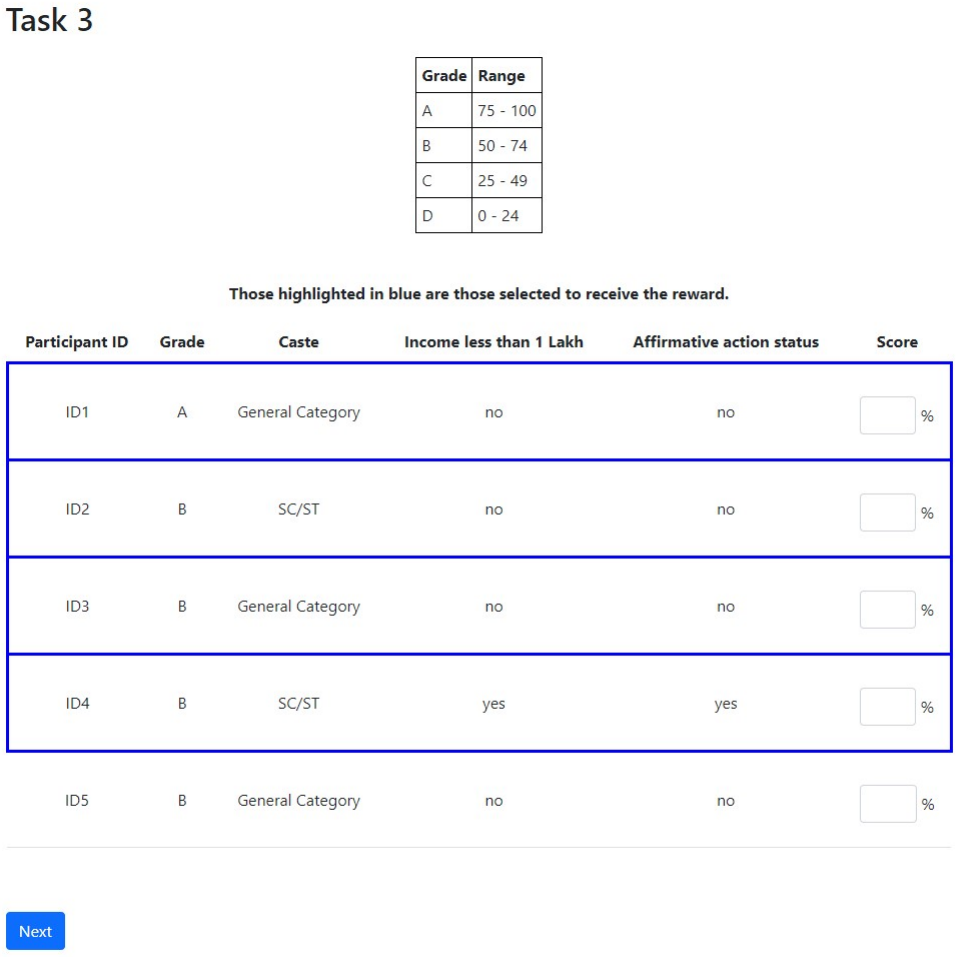


Figure A26: Score guessing: Group 3 - AA-Caste

Notes. Experimental screen shown to evaluators.

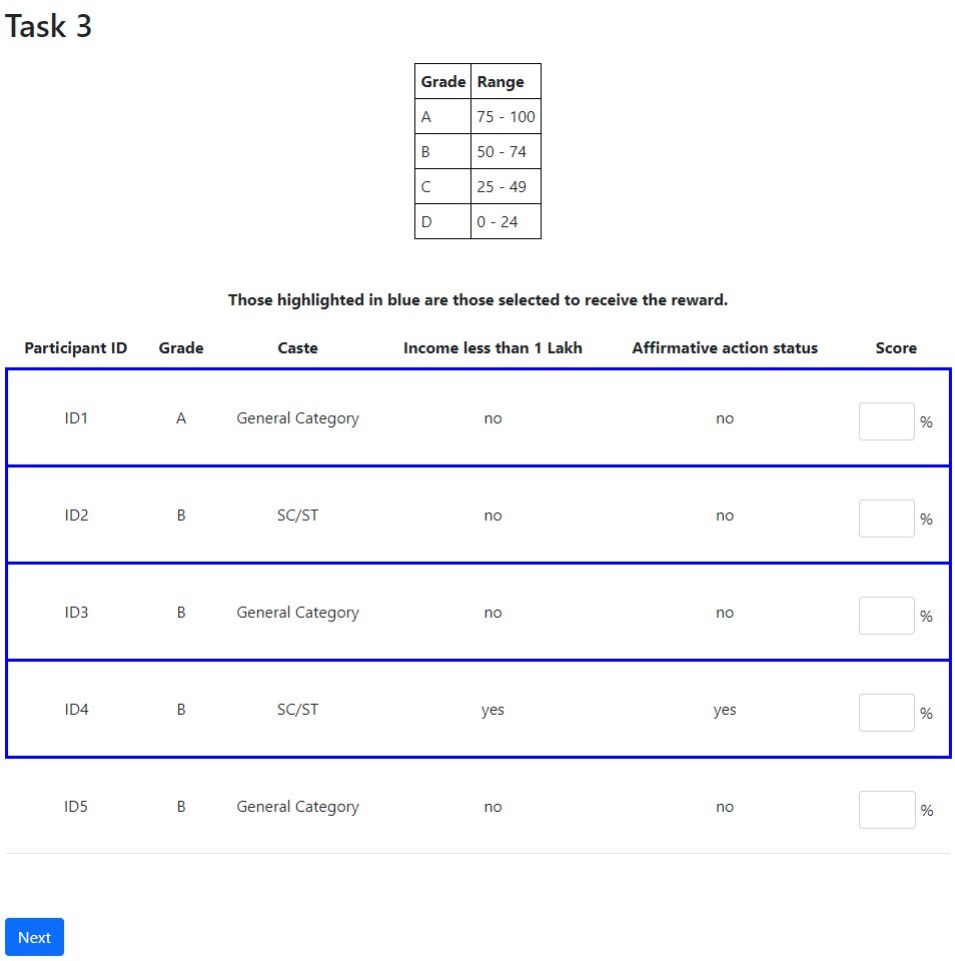


Figure A27: Score guessing: Group 3 - AA-Income

Notes. Experimental screen shown to evaluators.

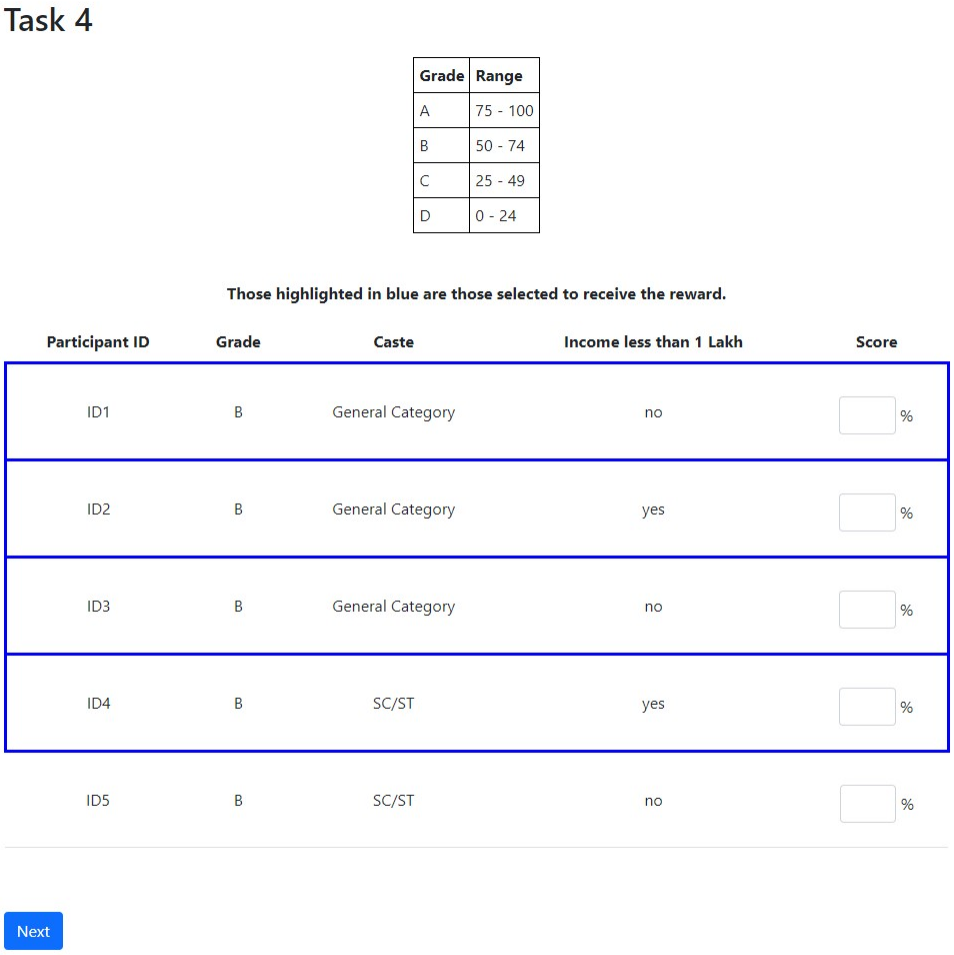


Figure A28: Score guessing: Group 4 - Control

Notes. Experimental screen shown to evaluators.

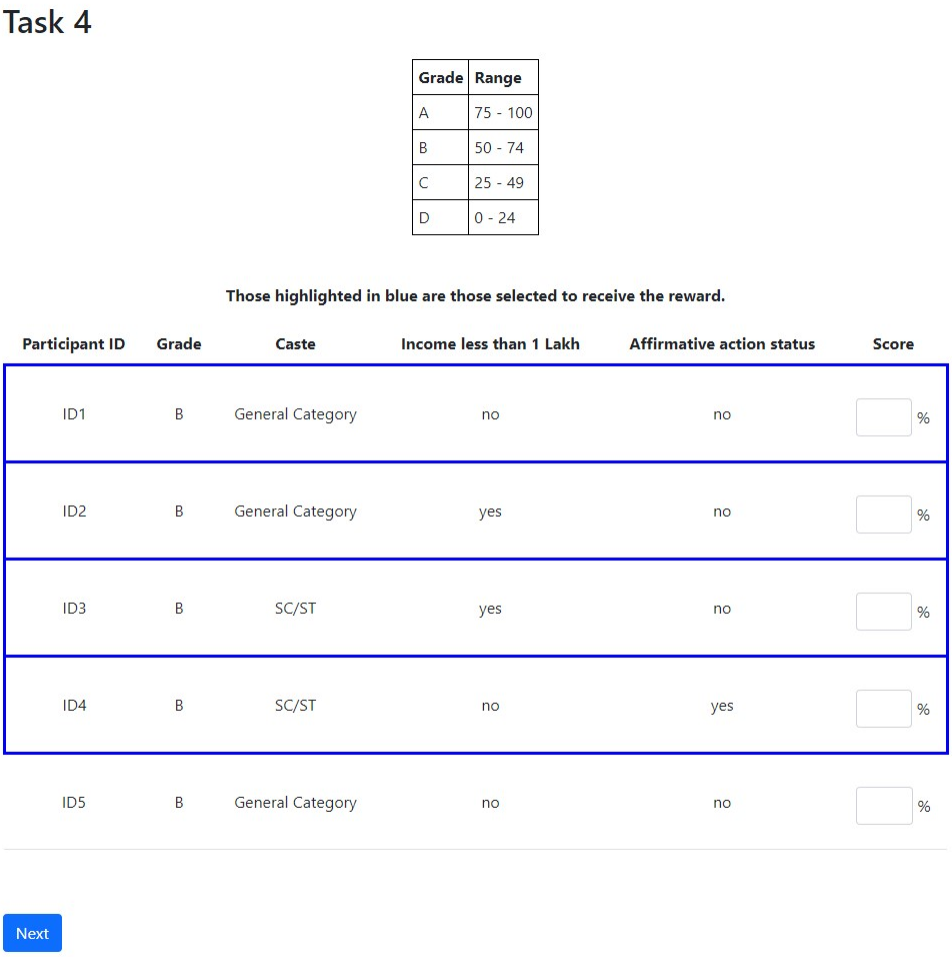
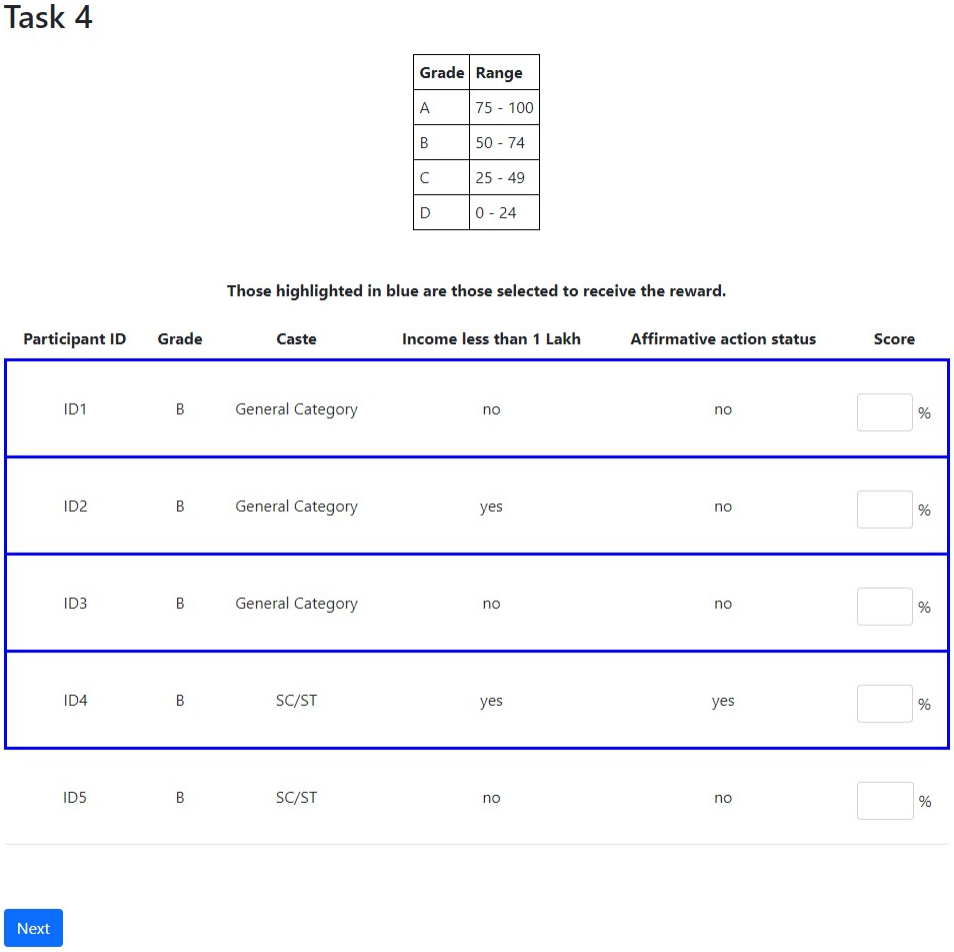


Figure A29: Score guessing: Group 4 - AA-Caste

Notes. Experimental screen shown to evaluators.



Task 4

Grade	Range
A	75 - 100
B	50 - 74
C	25 - 49
D	0 - 24

Those highlighted in blue are those selected to receive the reward.

Participant ID	Grade	Caste	Income less than 1 Lakh	Affirmative action status	Score
ID1	B	General Category	no	no	%
ID2	B	General Category	yes	no	%
ID3	B	General Category	no	no	%
ID4	B	SC/ST	yes	yes	%
ID5	B	SC/ST	no	no	%

Next

Figure A30: Score guessing: Group 4 - AA-Income

Notes. Experimental screen shown to evaluators.

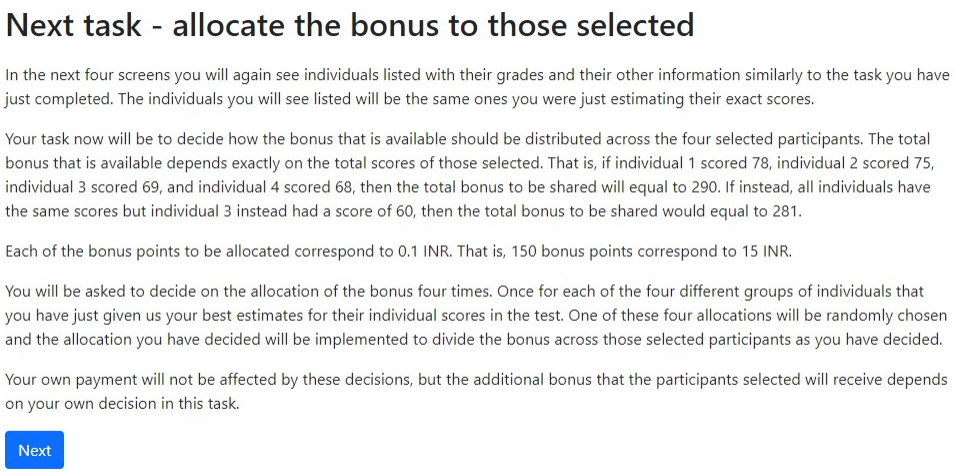


Figure A31: Bonus allocation instructions

Notes. Experimental screen shown to evaluators.

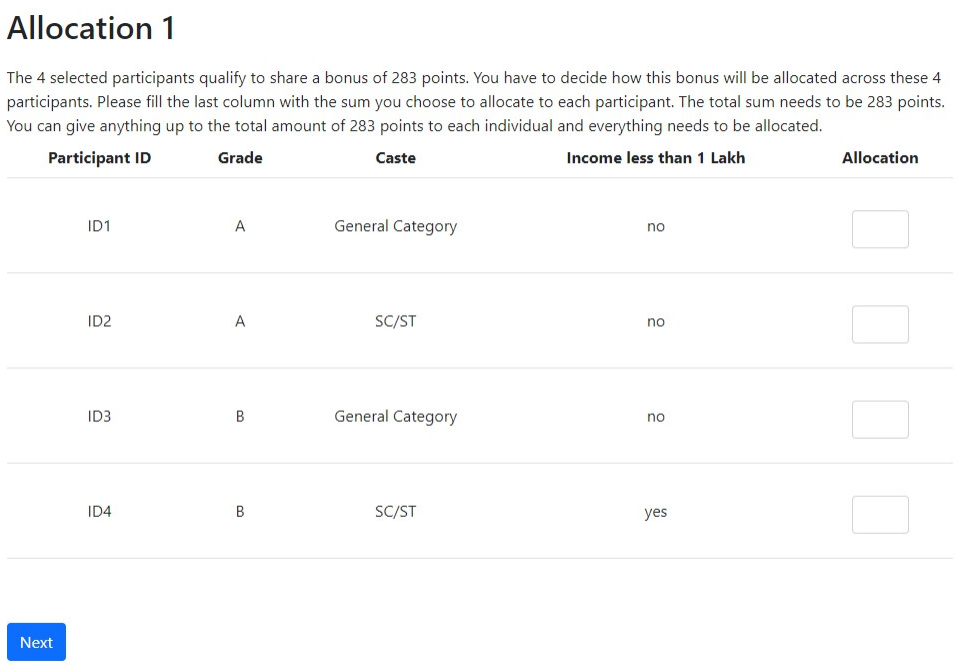


Figure A32: Allocation: Group 1 - Control

Notes. Experimental screen shown to evaluators.

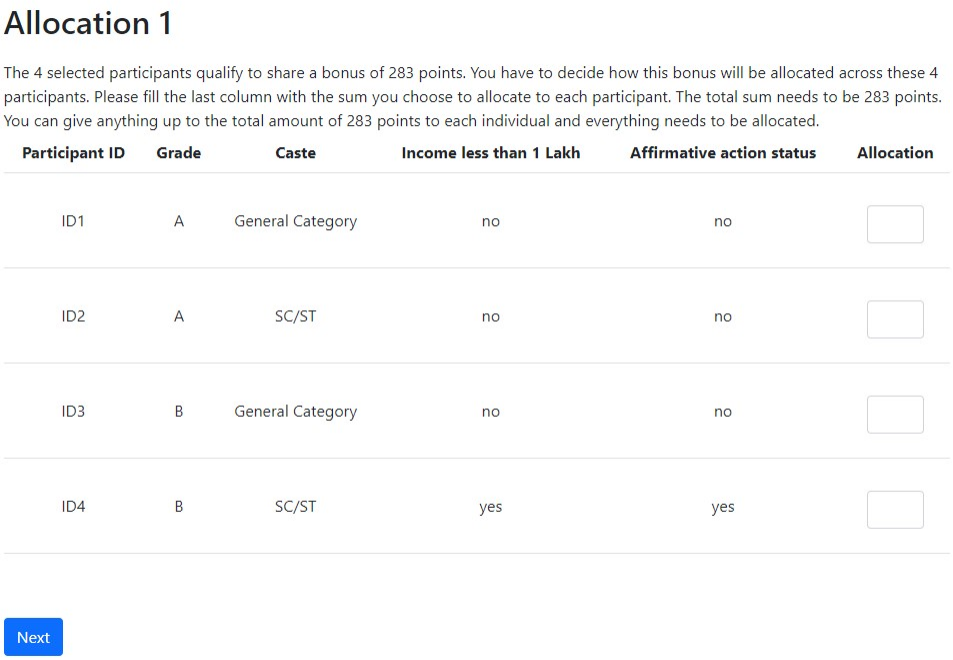


Figure A33: Allocation: Group 1 - AA-Caste

Notes. Experimental screen shown to evaluators.

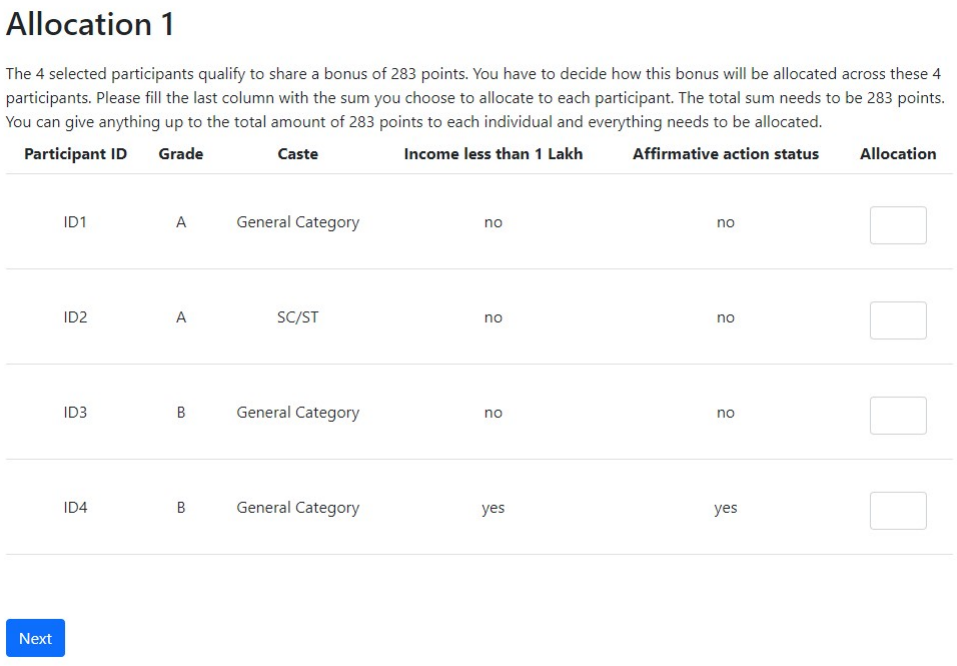


Figure A34: Allocation: Group 1 - AA-Income

Notes. Experimental screen shown to evaluators.

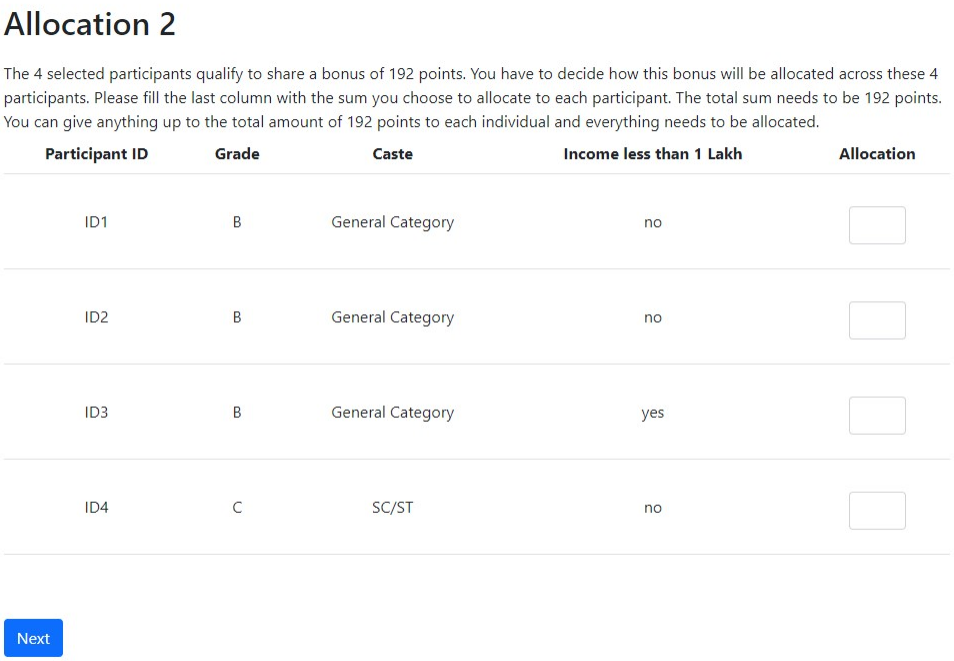


Figure A35: Allocation: Group 2 - Control

Notes. Experimental screen shown to evaluators.

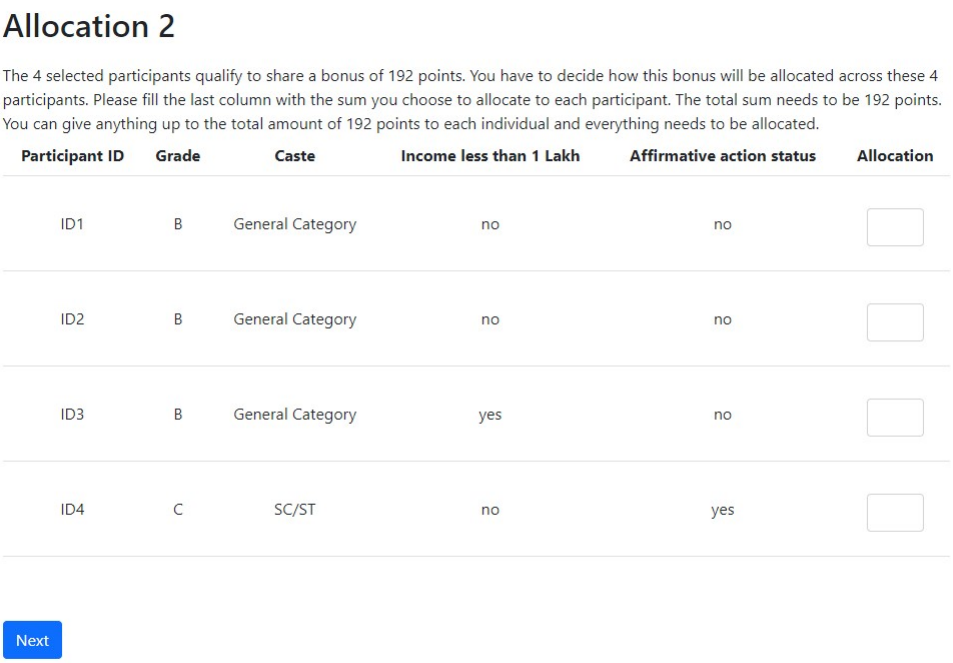


Figure A36: Allocation: Group 2 - AA-Caste

Notes. Experimental screen shown to evaluators.

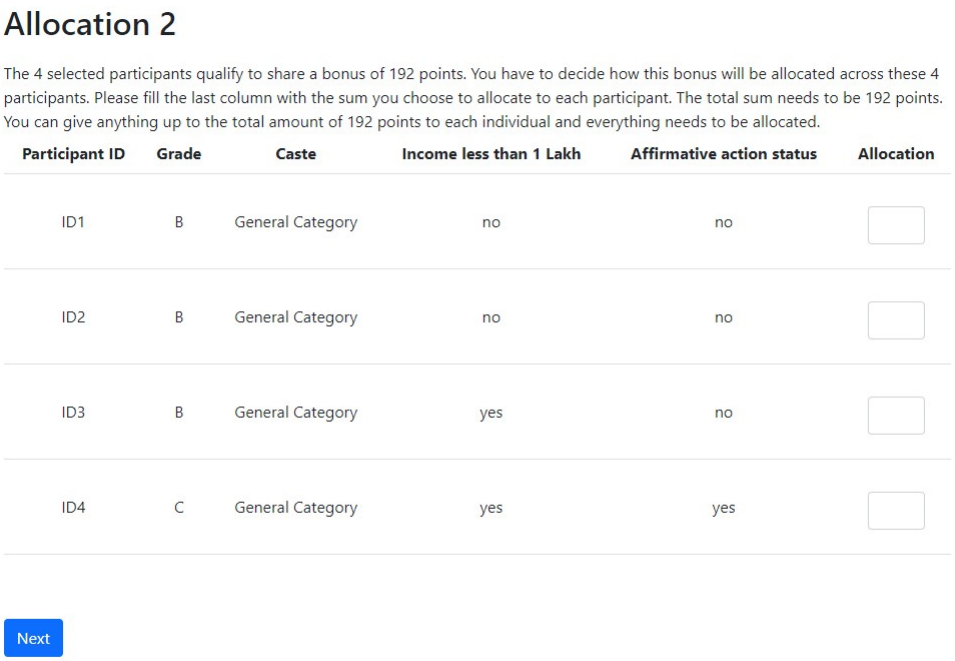


Figure A37: Allocation: Group 2 - AA-Income

Notes. Experimental screen shown to evaluators.

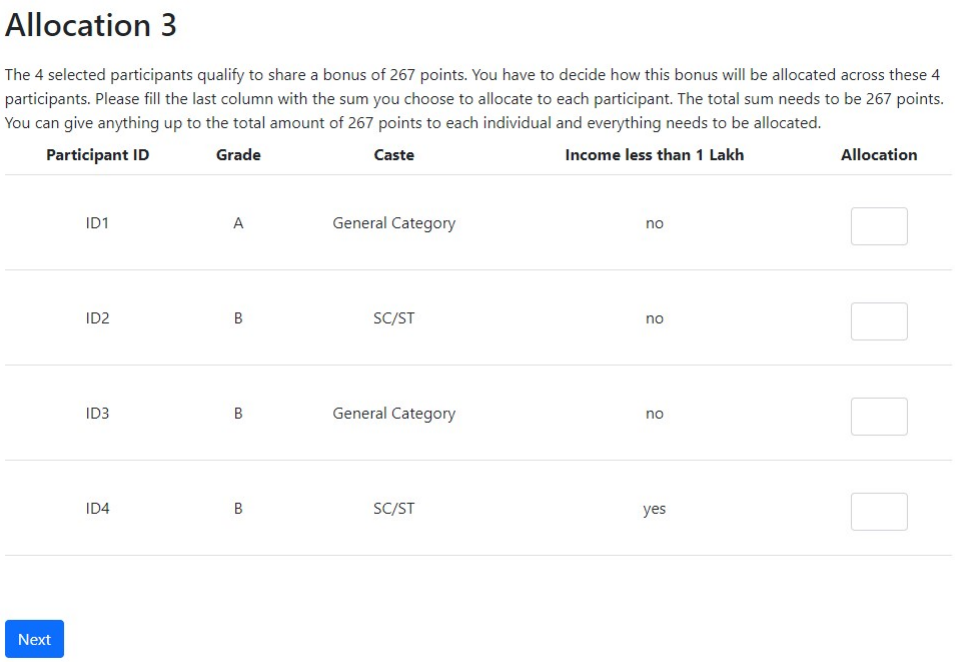


Figure A38: Allocation: Group 3 - Control

Notes. Experimental screen shown to evaluators.

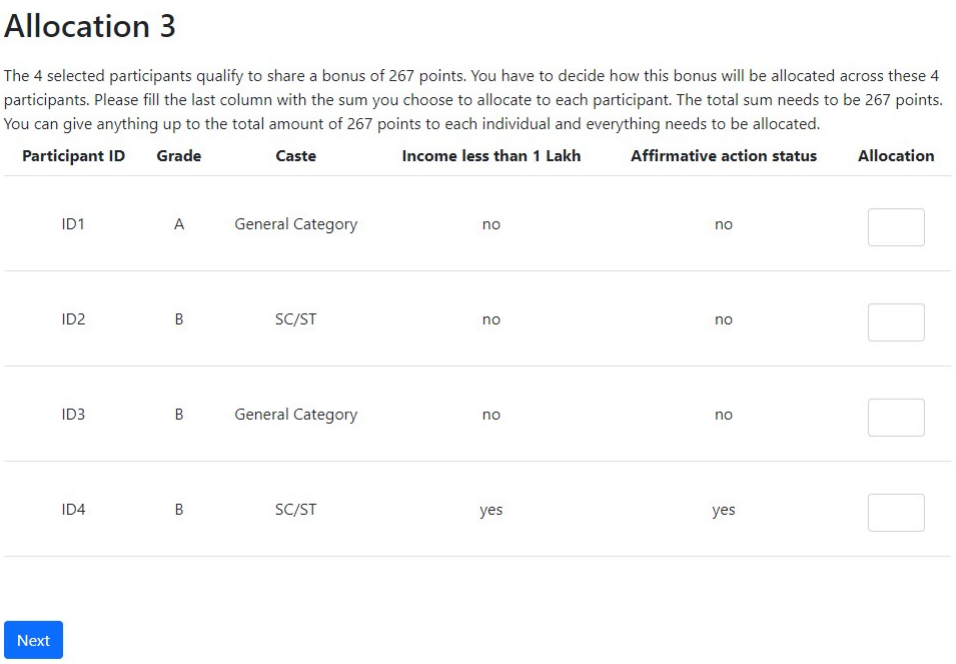


Figure A39: Allocation: Group 3 - AA-Caste

Notes. Experimental screen shown to evaluators.

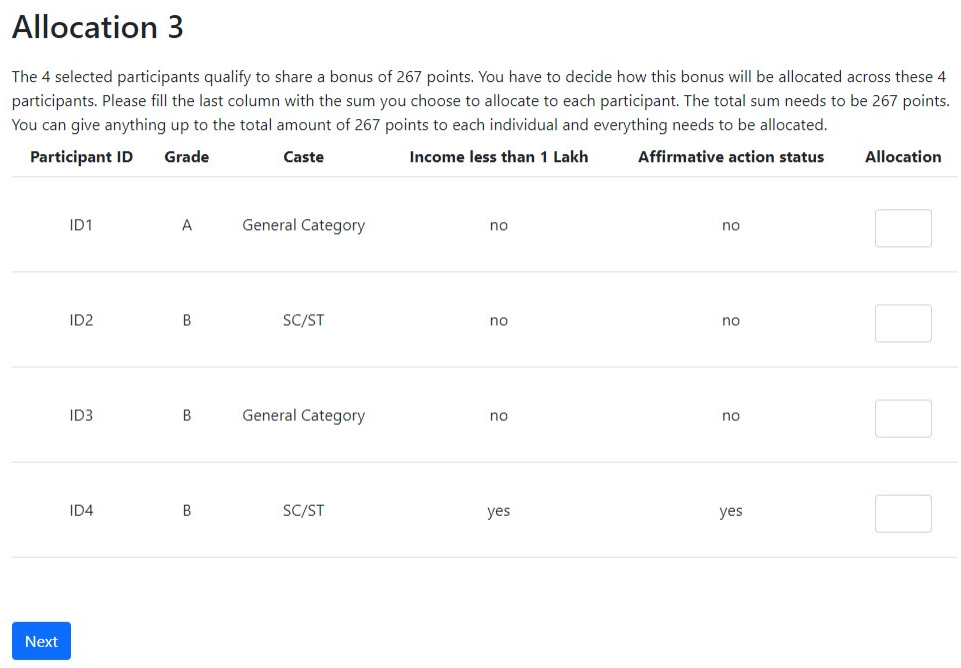


Figure A40: Allocation: Group 3 - AA-Income

Notes. Experimental screen shown to evaluators.

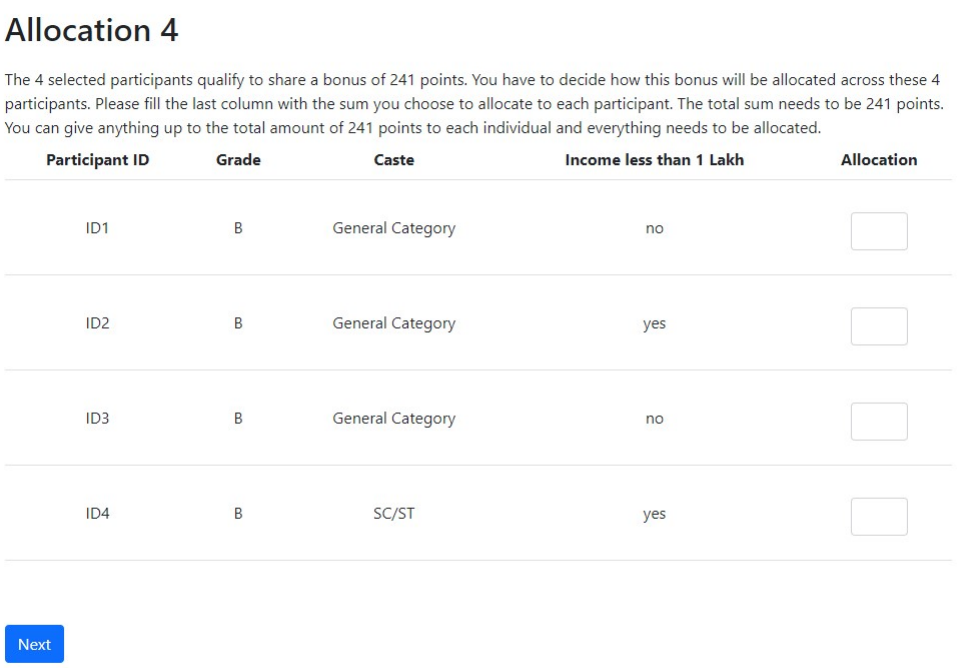


Figure A41: Allocation: Group 4 - Control

Notes. Experimental screen shown to evaluators.

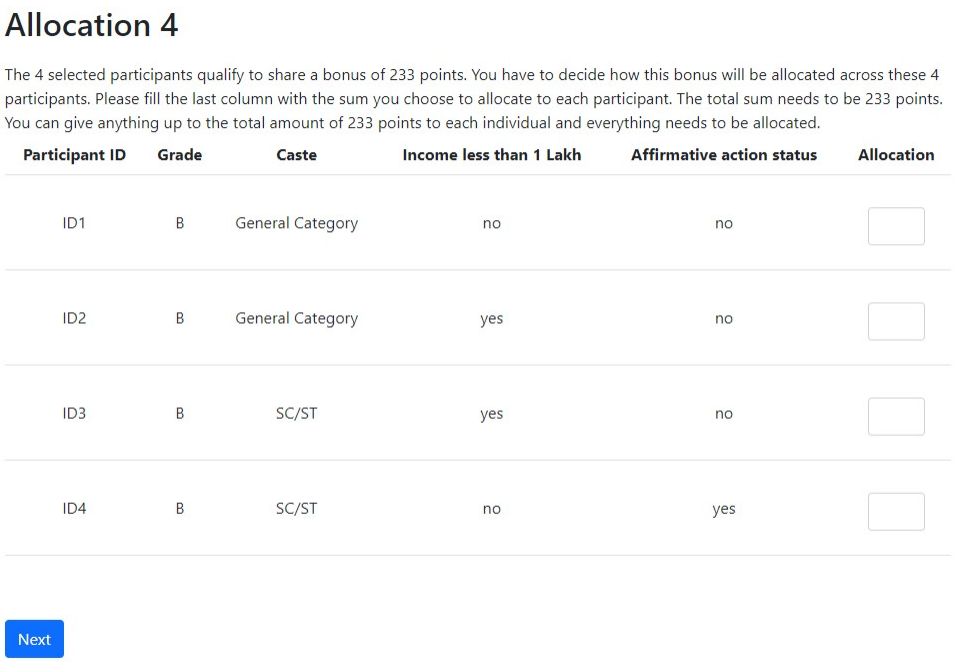


Figure A42: Allocation: Group 4 - AA-Caste

Notes. Experimental screen shown to evaluators.

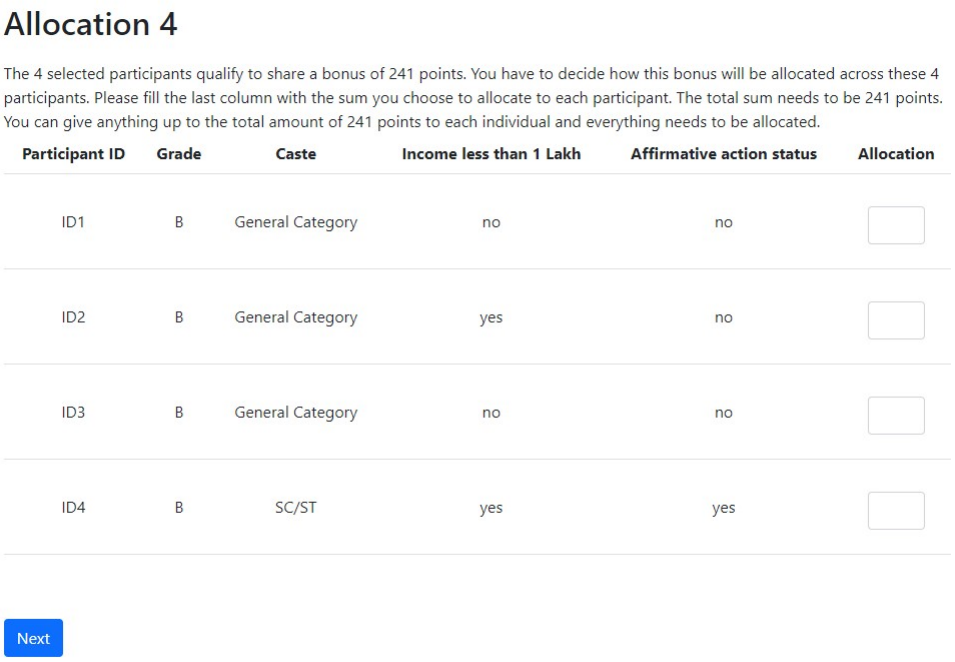


Figure A43: Allocation: Group 4 - AA-Income

Notes. Experimental screen shown to evaluators.

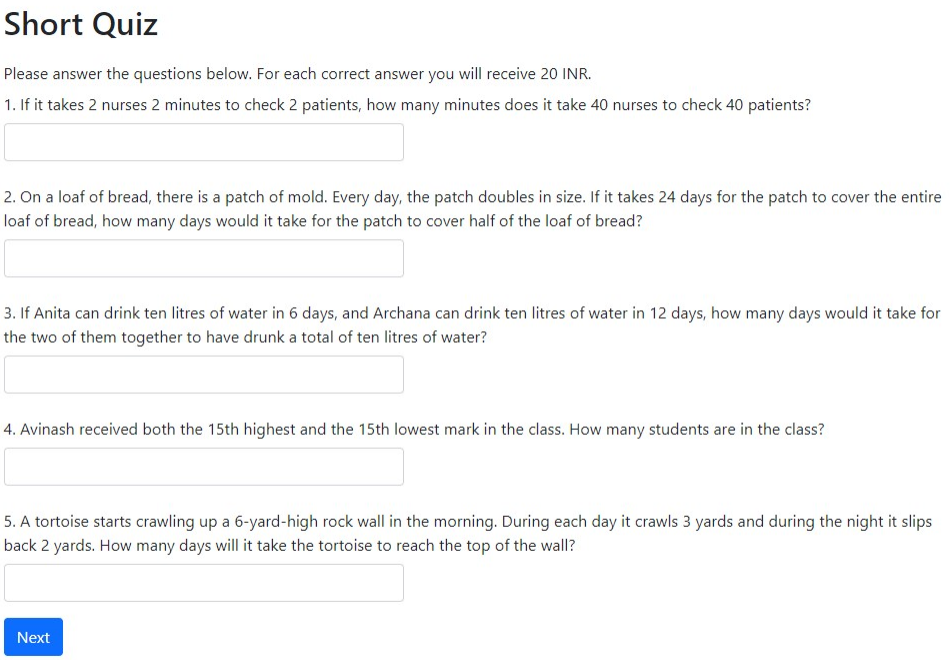


Figure A44: CRT

Notes. Experimental screen shown to evaluators.

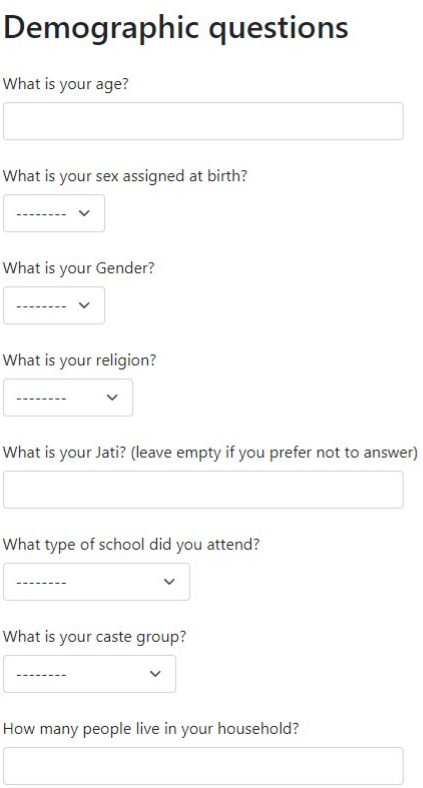


Figure A45: Demographics Questionnaire - Part 1

Notes. Experimental screen shown to evaluators.

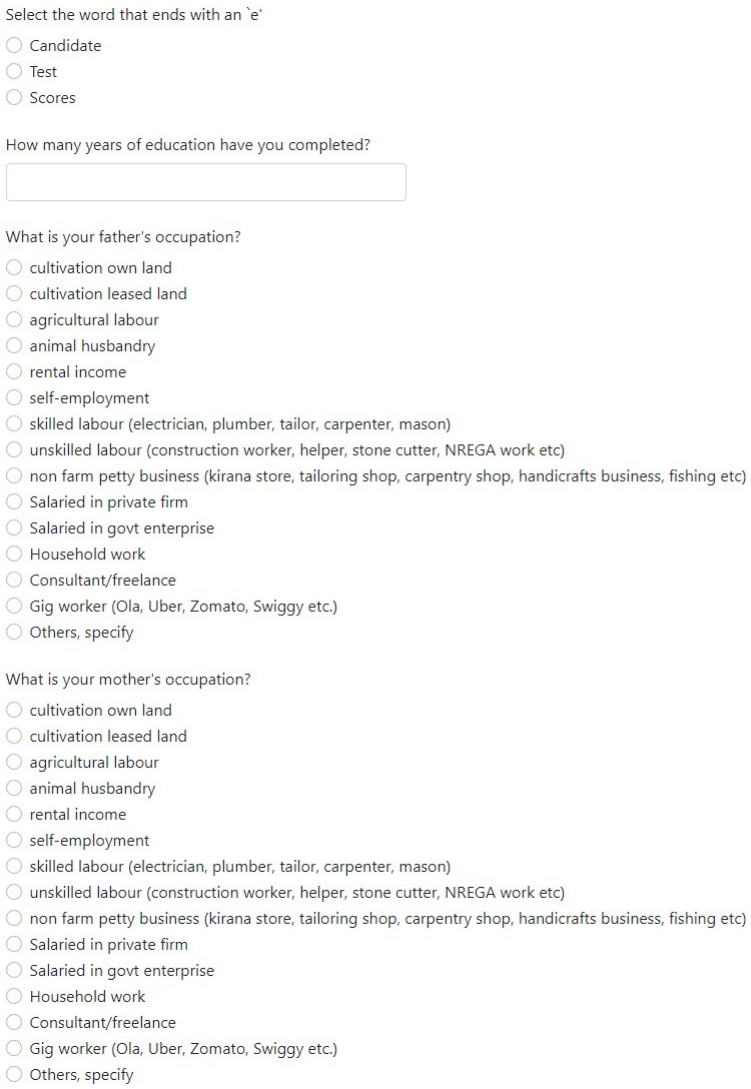


Figure A46: Demographics Questionnaire - Part 2

Notes. Experimental screen shown to evaluators.

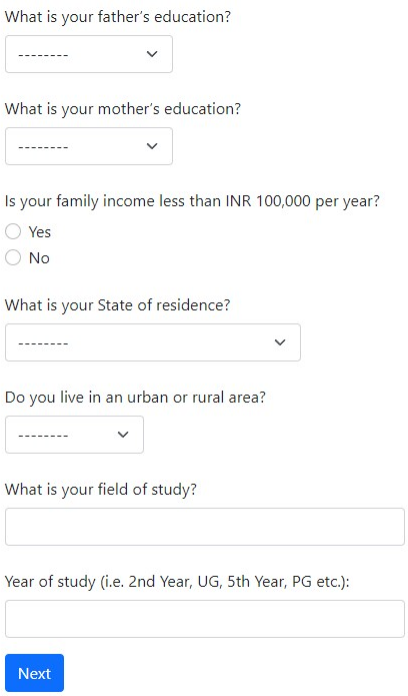


Figure A47: Demographics Questionnaire - Part 3

Notes. Experimental screen shown to evaluators.

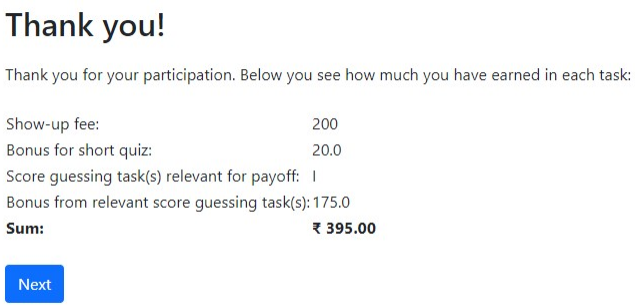


Figure A48: appendix˙exhb Screen

Notes. Experimental screen shown to evaluators.

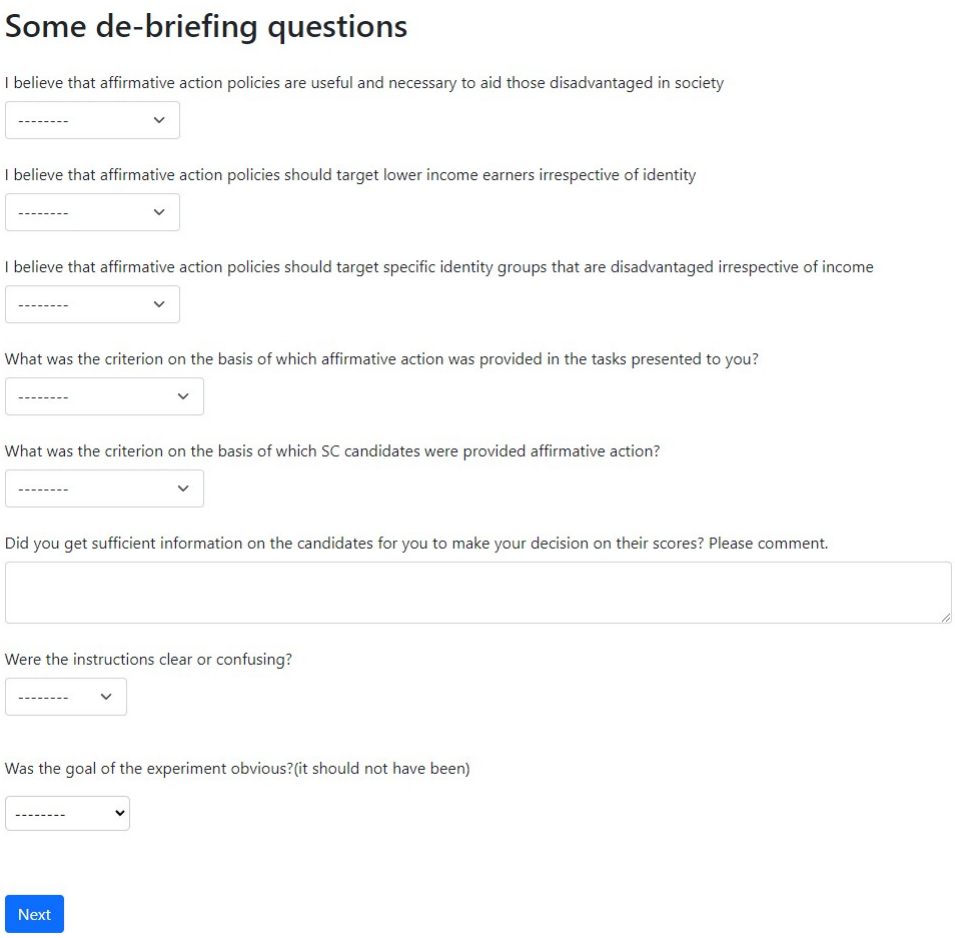


Figure A49: De-briefing questions

Notes. Experimental screen shown to evaluators.