Does negative marking lead to gendered exam taking behavior: experimental evidence from India*

Swati Sharma†
Nandana Sengupta‡
Sumitava Mukherjee§

December 7, 2024

Abstract

Negative marking is used globally in competitive examinations to deter random guessing and capture true academic rigor of students. However, it may create gendered outcomes due to systematic behavioral differences, such as lower confidence or higher loss aversion. We mimicked a competitive exam set up in a behavioral experiment with 275 B.Tech. students from high-stake STEM Institutes in India where participants undertook a time pressured MCQ with participants randomly assigned to sessions with and without negative marking. We find a fall of approximately 10% in number of attempts and of 7% in number of correct answers for females in the negative marking group with no impact on their conditional accuracy rates. This has a direct implication on their total scores and ranks. Intriguingly, our results are driven by the quantitative section which was technically less advanced and familiar as opposed to the logical reasoning section. We attempt to underpin the mechanism for our results by looking at loss aversion, perceptions about oneself and others, as well as self-confidence in a competitive environment. Our study is the first to causally estimate the impact of negative-marking on test outcomes in a developing country with policy implications for the screening processes used by STEM institutes across the globe.

*Preliminary draft. Do not cite. We thank CoImpact for the financial support. Ethics approval for conducting the study obtained from Institute Ethics Committee of Indian Institute of Technology Delhi, (IEC-IITD) with approval number 2021/P092. Study registered with AEA-RCT Registry (RCT ID AEARCTR-0012042) with the same title.

†Indian Institute of Technology, Delhi (swati.sharma@hss.iitd.ac.in)
‡Indian Institute of Technology, Delhi (nandana@sopp.iitd.ac.in)
§Indian Institute of Technology, Delhi (sm1@hss.iitd.ac.in)

KEYWORDS: MCQ, negative marking, loss aversion, gender, STEM institute, India, Experiment

JEL CLASSIFICATION: C91, D91, I23, I24

1 Introduction

Negative marking is used globally in Multiple Choice Questions (MCQ) based competitive exams to deter random guessing and to capture true academic rigor of the applicants. Theoretically, negative marking attaches loss/risk if a candidate is uncertain about the correct answer. Simulation studies by Espinosa and Gardeazabal (2010), Karandikar (2010) show that a candidate with a higher degree of risk aversion will end up with lower scores even with commonly accepted expected score equalizing penalty rule of $\frac{1}{k-1}$, where *'k'* is number of options, for a given level of knowledge. Moreover, penalty based scoring rules have been found to be inefficient in capturing partial knowledge (Bliss (1980)). Since females have been identified to be more risk averse, loss averse and shy away from mixed gender competitive settings, and less confident in different contexts, a penalty for an incorrect answer may lead to suboptimal number of answers, thereby affecting their selection (Byrnes et al.(1999), Croson & Gneezy (2009), Eckel & Grossman (2008), Dawson (2023), Niederle & Vesterlund (2007)). Therefore, it becomes important to examine the impact of negative marking on performance (scores) by gender.

Another stylized fact is that even though girls consistently outperform boys at the subjective/open ended examinations both at school and college (university) level (.i.e, the evaluations preceding and following competitive exams), the trends totally reverses when we look at female's performance in most of the high stake competitive exams (Graetz and Karimi (2022)). Moreover, gender gaps in average education attainment have been closing but with little or no impact on gaps in disciplines that rely heavily on Negative marking-MCQ based competitive exams for screening aspirants. Recent policy actions to address gender gaps in competitive examinations like removal of penalties in US-SAT (2016), Advanced Placement Program in US

(2012), Ghent University exams in Europe, and University Selection Tests, Chile (2019) have led to a significant increase in female representation in the respective contexts (Coffman and Klinowski (2020), Graetz and Karimi (2022)). In light of this evidence, discourse around gendered nature of different elements of competitive exam has been gaining momentum.

Literature predominantly comes from field or natural experiments and establishes negative relationship between grading patterns and performance of females (Burns et al. (2012) from South Africa, Espinosa and Gardeazabal (2013) from Spain, Pekkarinen (2015) from Finland, Foster (2016) from UK, Coffman and Klinowski (2020) from Chile, Saygin and Atwater (2021) from Turkey). However, some of the observational studies do indicate an increase in precision or strike rates giving null to mixed results by ability type and gender (Senela et al. (2013) from Turkey, Fisk (2018) from USA, Akyol et al. (2022) from Turkey, Funk (2023) from Spain). Notably, evidence and policy discourse has overwhelmingly been limited to the developed countries or global North.[1] Moreover, these studies are mostly natural or field experiments or observational suffering from self-selection and have too many confounding channels to causally pin down the impact of negative marking.

In this study we address this gap by randomly assigning participants to the penalty (negative marking in scoring rule) and no penalty (no negative marking in scoring rule) treatment in a real effort based task where subject's performance and grading scheme determined their earnings. The only other study that has looked at causal impact of penalty along with framing eliciting different stakes on outcomes is by Baldiga (2014) where 406 Harvard students were tested using twenty History SAT II questions. Baldiga (2014) identifies systematic difference in risk aversion by gender that

[1]The only exception being the study from South Africa, where authors varied marking scheme through out the micro economics course and find women to be more negatively responsive with an increase in the magnitude of the penalty.

partially explains skipping behavior of women and thus lower scores with penalty and high stake framing. We improve upon this design by keeping stakes uniform across the treatment and the control groups and only vary marking scheme. We also test two types of subject matter- (i) Simple Arithmetic and (ii) Logical Reasoning. The choice of the two subjects comes from the overwhelming stereotype literature establishing that males enjoy or have advantage in mathematical problem solving and females have advantage or likely to succeed in language/literature related questions and these stereotypes while absent till middle school become stronger in adulthood.[2] Our experimental design is also in line with studies like Riener and Wagner (2017), Saygin and Atwater (2021) that find impact of negative marking to be more pronounced for difficult questions and quantitative questions, respectively. Further, following **???**, we also look at measures of loss aversion in the post task questionnaire. We also attempt to capture the beliefs regarding own performance, confidence and self perception in a competitive setting and whether these have differential bearing in the presence of negative-marking.

Our study is not only the first to isolate the impact of negative marking on outcomes by gender but also extends existing literature to a developing country context where social conditioning, gender norms and modern institutions may play out differently (Jayachandran (2018), Munshi & Rosenzweig (2004)). We focus on two high stake Indian STEM Institutes of Eminence that admit students through negative marking MCQ based competitive exam.[3] Interestingly, gender gaps in STEM at high school and regional engineering colleges in India are narrow but same is not true of high stake nationwide engineering institutes (AISHE, Klasen and Sahoo (2020)).[4]

[2]See Helene et al (2023), K. Morrissey, et al (2019), Hyde et al. (2008), Hyde & Mertz (2009), Farenga & Joyce (1999), Eccles & Jacobs, (1986); Fennema & Sherman (1978).

[3]Penalty in entrance exams of IoEs (called IIT Mains and IIT Advanced) ranges from a $\frac{1}{3}$ to $\frac{1}{2}$ per incorrect answer which gets subtracted from total score.

[4]Females comprised 8-12% of B.Tech degree batch size across the high stake institutes till 2017. After the

Notably, on one hand these STEM institutes have a history of low and stagnant female presence especially in undergraduate degrees and on another females entering these institutes exit with better ranks as compared to their entering ranks (Sharma & Sengupta (2024)). Further, gender gaps get reversed if we look at natural sciences at university level that mostly use higher secondary exam scores as selection criterion.[5] Thus, our broader question of interest is whether negative marking (penalty based scoring rule) has the potential to explain a part of skewed entry of females in STEM courses.

We used a two by two between subjects design in a behavioral lab experiment set up and varied scoring (marking) scheme on a mixed effort quiz (having two sections – i. Simple Arithmetic, and ii. Logical Reasoning), where scores determined participant's monetary payoffs. Our study consisted of 275 B.Tech students from two of the high stake STEM Institutes of Eminence (IoEs) in India. 146 participants were subjected to the negative marking based scoring rule, i.e., penalty treatment (+20 points per correct answers, -10 points per incorrect answers, 0 points for skipping) and 129 to the no-penalty treatment (+20 points per correct answers, 0 points per incorrect answers, 0 points for skipping) across seven sessions.

We find that negative marking (-$\frac{1}{2}$ per incorrect answer as opposed to no negative marking scheme) lead to lower net scores, worse off ranks, lower attempts, lowers unconditional accuracy rates (i.e., number of correct answers out of 40 assigned ques-

introduction of Supernumerary Seats Schemes in 2018, the proportion of females in same degrees reached 20% by 2020. Trend has been more or less stagnant ever since (Sengupta et al (2024)).

[5]For example, some of the Central-Public Universities like Delhi University (largest Centralized University in India), Aligarh Muslim University, and State-Public universities like Calcutta University, Jadavpur University, Madras University for admission to natural sciences courses like B.Sc. in Physics, Mathematics, Botany, Computer Sciences, etc. use higher secondary scores alone. Entrance exams are a norm for post-graduation admissions in almost all the education set ups in India, whereas professional courses use entrance exams for admission processes even at graduation levels, as mentioned earlier. Delhi University introduced CUET in 2021 that uses MCQ-negative marking (5 points per correct answer, -1 per wrong answer) for screening eligible students. A drop of 25% in female enrollment rates was observed in the same year with trend being monotonically increasing by 2023 (`https://indianexpress.com/article/education/after-cuet-du-admissions-drop-25-enrollment-of-girls-sees-bulk-of-decrease-8424421/`).

tions), but increase conditional accuracy rates (i.e., number of correct answers out of attempted questions). However, impact of negative marking is significantly more pronounced for females in penalty treatment – a fall of 10% (4 out of 40 questions) in number of attempts and 7% (3 out of 40 questions) in number of correct answers. Females' unconditional accuracy rates fall by 6.3 p.p but we find no impact on conditional accuracy rates. The clear implication of penalty based scoring rule is lower scores and ranks of females due to fall in number of attempts with no impact on precision. Intriguingly, overall results are driven by differences in the females' performance in simple arithmetic section which was the less advanced and more familiar of the two.

Our results hold with institute-branch fixed effects, question fixed effects and a host of individual characteristics and general academic performance. We discuss some of the plausible underlying mechanisms and provide evidence for the same. Our analysis concludes that penalty has a stronger negative impact on females due to fall in number of attempts coupled with no impact on their conditional accuracy rates. Notably, the results are driven by familiar and less advanced section for which many of these students undertake coaching while preparing entrance exam.[6] Therefore, what is being tested matters for inclusivity. An examiner, thus, needs to carefully design the content and scoring scheme and be cognizant of their implications for the historically disadvantaged groups.

One must note two points here – one being that our results are internally valid and come from extremely high ability students with narrower gender gaps in the average abilities. We expect both males and females to have similar rigor in mathematics

[6]Coaching culture and gendered access to coaching centers for preparation of entrance exams of STEM IoEs is well known in Indian context. Students at these coaching centers are taught tricks like answering through elimination/backward engineering in face of uncertainty which may lead to better/calculated guessing behavior.

because these students have cleared series of highly competitive quantitative exam to make it to these institutes and also because we find no differential impact of gender per se. Thus, ability difference by gender explanation seems less plausible. Second point is that we find gendered effects to be larger in magnitude in the relatively low ranked (yet high stake) institute in our sample. Thus, our estimates are plausibly the lower bounds on gendered impact of the penalty in scoring rules and these impacts may be even stronger with other existing gaps like ability or inequalities like investment in education, etc. In light of the intense competition that exists around these screening exams, a difference of even half a point (through let's say getting even one more question correct) may alter one's entire career trajectory.

While we focus on gender and high stake STEM institutes to study impact of negative marking, the context is not limited to just STEM as many exams in India and over the world use negative marking in the screening process. We describe some of these in the section 2, followed by the description of the experiment in the section 3. We describe our data in section 4 and discuss empirical methodology and results. section 5 discusses results and mechanisms and section 6 concludes.

2 Context and Background

2.1 Negative marking and competitive exams in India

Negative marking is used commonly in MCQ based competitive exams to deter random guessing. Some examples of nationwide entrance exams with negative marking in India are NEET for entrance in medical colleges, CAT for high stake management institutes, SSC CGL exam for government jobs and IBPS PO Exam for a career in public banking sector and so on. Negative marking in these exams is around $\frac{1}{k-1}$, where *'k'* are number of options. Interestingly, negative marking ranges from $\frac{1}{3}$ to $\frac{1}{2}$ per incorrect answer in the entrance of high stake STEM institutes (stage

I named IIT Mains and stage II named as IIT Advanced) which by far is the strictest in Indian examination context. Not only fewer women register for IIT Mains and IIT Advanced, fewer qualify as compared to men.[7]

While discourse on negative marking and gendered outcomes have led to policy changes at national level examinations in countries like U.S., Belgium, Chile to see more inclusive outcomes, India has been moving towards penalty based scoring in more and more exams. A recent example is introduction of CUET in 2021 in Delhi University which is one the largest and most prestigious Central university in India. CUET uses MCQ-negative marking (5 points per correct answer, -1 per wrong answer) for screening eligible students. Delhi University saw a drop of 25% in female enrollment rates in the year of the introduction with trend being monotonically increasing by 2023.[8]

The advantages of negative marking MCQ type based screening process are of course many – easier to conduct, evaluate and cost effective, especially in a developing country where millions of students apply for national level examinations and institutes. However, it is also important to keep in mind that the screening process should not bias the results on the basis of how knowledge is tested.

2.2 High stake STEM institutes and gender disparity

Given the importance attached to a professional tertiary education like B.Tech (engineering), it's potential to define one's career trajectory and higher rates of returns

[7]NEET like IIT Advanced is a highly prestigious competitive exam in India that opens doors for a medical profession in high stake medical institutes for top rankers. Contrary to IIT Mains and Advanced, NEET has a history of higher proportion of females applying and qualifying as compared to males. While NEET tests questions from Physics, Chemistry and Biology, IIT Mains and Advanced tests Physics, Chemistry and Mathematics. Our paper focuses on the impact of negative marking on test takers performance in general. It is beyond the scope of our paper to outline the historic factors resulting in different gender representation in the two examinations. However, our results do provide some suggestive explanations as discussed in the section 5

[8]Source:`https://indianexpress.com/article/education/after-cuet-du-admissions-drop-25-\enrollment-of-girls-sees-bulk-of-decrease-8424421/`

as compared to other disciplines (Jain et al. (2021)), our study draws sample from two high stake STEM Institutes in India. Further, studies also show that females pursuing STEM degrees have lower probabilities of dropping out of labor market and have better labor market outcomes that makes our context even more important in the light of low and stagnant female labor force participation in India and occupational segregation (Şahoo and Klasen (2018)).

While gender gaps are considerably narrower when we look at female presence in Science stream at the high school level, degree programs in universities, natural sciences and regional B.Tech colleges vis-e-vis high stake STEM institutes like IIT, IIITs, IISc or IISERs (AISHE reports). For example, 49% of the students enrolled in PCM at high school level are girls, but proportion of girls was around 8-10% females in B.Tech batches across IITs before 2017. It has capped now at 20%, post introduction of an affirmative action scheme called Supernumerary Seats Schemes (for details see Sengupta et. al. (2024)). One of the key difference in the screening process of latter is heavy reliance on highly competitive negative marking based exam consisting of questions from Physics, Chemistry and Mathematics.[9][10]

The two high stake STEM institutes in our sample follow same screening process – common admission and counselling process conducted by JOSAA (Joint Seat Allocation Authority). Both the institutes have been identified as Institutes of Eminence (IoE) by Indian government and are in top 50 according to the NIRF rankings (National Institute Ranking Framework governed by Ministry of Education, Government of India). Even though both are high stake STEM institutions there is a considerable gap in the ranking and prestige attached to each. For reference purposes, on the basis

[9]Students are also required to have a score of 65% to 75% at high school (XII) level as a qualifying condition.

[10]We do not contest that negative making explains skewed gender ratio in selection of B.Tech degree programs entirely but we study if negative marking has a role to play in skewed gender distribution of qualifying candidates.

of NIRF ranking, we will address both institutes as Institute 1 and Institute 2 with Institute 1 being relatively better ranked, larger in size and older. Nevertheless, both have similar gender environment in terms of gender representation in B.Tech courses.

Our key interest is to study if negative marking based scoring rule leads to capturing of academic rigor of girls inefficiently. It is in light of the research work by Sharma & Sengupta (2024) based on same institutes which shows that even though girls enter with worse ranks on an average, they exit with much better ranks showing remarkable *'catch up'*.[11] It is also in line with global trend of women outperforming men when it comes to GPAs and other long term metrics as opposed to performance in competitive exams. Graetz and Karimi (2022) argue that subjective school-college level exams measure traits like 'conscientiousness', effort and motivation at which girls are better unlike competitive examinations that might be affected by traits at which girls are disadvantaged.

3 Experimental design

3.1 Subjects and recruitment

Figure A.1 gives a broad description of the experimental design flow.[12] The entire experiment with in an institute from the stage of eliciting interest to final payments took about a month.[13] It started with sending out of institute-wide email a week prior to the effort based task with broad details regarding potential payoffs and gross time required for a "non academic fun quiz". Details of call are in Appendix B. Students were required to register first and registration process closed three days prior to the

[11] The same rationale led to the introduction of Supernumerary Seats Scheme.

[12] Ethics approval for conducting the study obtained from Institute Ethics Committee of Indian Institute of Technology Delhi, (IEC-IITD) with approval number 2021/P092. Study registered with AEA-RCT Registry (RCT ID AEARCTR-0012042) with the same title.

[13] Study was approved by Institute Ethics Committee of Indian Institute of Technology Delhi, (IEC-IITD) with approval number 2021/P092. Experiment was registered with AEA-RCT Registry (RCT ID AEARCTR-0012042) with the same title before conducting the pilot.

task. Experiment was conducted in October 2023 and April 2024 in Institute 1 and Institute 2, respectively.[14]

We called about 50% of students keeping in mind the compliance rates from past experience of conducting experiments in the same institutes (Sengupta 2023). Our final sample consisted of 275 students.

Participants only knew that they would be participating in a non-course related quiz (call specified "fun non-academic quiz", see Appendix B) to be held on computers. All our participants were familiar with this type of exam taking method and also with MCQs with penalty scoring as all of them have cleared IIT Mains and IIT Advanced.

An experimenter outside the parallel labs, assigned students randomly to the separate labs to keep gender ratio similar in all the rooms. As students entered their respective computer labs, they were required to submit their phones and proceed to their assigned seats. Students were seated with a gap of one system to inhibit cheating or communication. Every row had two volunteers patrolling to keep a check and assist participants. Each session and row had an equal mix of male and female volunteers. As soon as participants were seated they marked their attendance on the sheet that an another volunteer was in charge of getting signed. Once everyone was seated, head experimenter (female) in all the sessions read out instruction (given in Appendix C) loud and clear and encouraged participants to ask any clarificatory questions. We wanted to make sure that everyone understood the rules and payoff structure and thus it was unavoidable to conduct penalty and no-penalty sessions

[14]Approximately 860 students registered across the two institutes, of which (not correcting for double registrations), of which approximately 23% identified themselves as females.[15] In order to ensure empirically viable number of females, we over sampled girls by recruiting everyone for participation and randomly selecting boys for participation. All the selected students were informed via emails, followed by calls one day before the study and invitations were sent out on the day of the experiment through the community messaging applications such as WhatsApp groups. Those male students who were not selected for participation were sent a non-selection acknowledgement email a day prior to the experiment.

separately and hence treatment was at the session level.[16]

A session took about 45 minutes - 10-15 minutes for settling down, 10 minutes for instructions, 24 minutes - effort based task, 5-10 minutes for post-task questionnaire, 5-10 minutes for wrapping up.

Overall, subjects understood the effort based task and as instructed no body asked quiz-question specific doubts during the quiz. We describe the effort based task and payoff structure as follows.

3.2 Instrument design

3.2.1 Effort based task

To make the quiz engaging and ensuring earnest efforts, the quiz consisted of two sections - (i) Simple Arithmetic (SA), (ii) Logical reasoning (LR).[17] Since, our potential sample consisted of engineering students from academically rigorous institutes, we used real problems instead of trivial questions unlike competitive literature (Niederle & Vesterlund (2007)). It is also in line with findings of some of the field studies that show that scoring rule matters for difficult questions, i.e., questions that involve some aptitude, effort and thinking (Espinosa and Gardeazabal (2005), Espinosa and Gardeazabal (2010), Riener and Wagner (2017)).

Simple arithmetic questions ranged from mathematics curriculum up till higher secondary school level in India. Since everyone must score at least 65% marks in mathematics at the aforementioned level to be eligible to write the entrance exam, our design discounts the lack of knowledge or ability differences to have a systematic influence on a student's performance. Also, literature emphasizes that gender stereotypes regarding mathematics appear only in adulthood, not at school level.[18]

[16]Details on protocol and instruction given in Appendix C.

[17]Types of questions and rigor levels were tested during pre-pilot (held in April 2023) and pilot (held in September 2023) with students from courses which were not part of our potential sample before finalizing.

[18]See Helene et al (2023), K. Morrissey, et al (2019), Hyde et al. (2008), Hyde & Mertz (2009), Farenga & Joyce (1999), Eccles & Jacobs, (1986); Fennema & Sherman (1978).

Twenty questions in SA can be mapped into 4 broad topics - (i). BODMAS (ii.) Matrix algebra, (iii) Linear equation, (iv) Basic Probability. Further, five questions of each type had a mix of low effort and high effort in each type of questions [19]

Contrary to Simple Arithmetic section, Logical reasoning was more advanced and unfamiliar section (GRE-NET level).[20] Similar to the pattern in Simple Arithmetic section, this section also had four type of questions ((i) Multiple statements (ii) Statement and Conclusions (iii) Cause and effect (lengthier) and (iv) Cause and effect (shorter)) with half of the questions being low effort type requiring less reading/thinking time for answering.

Our quiz consisted of two sections due to multiple reasons...

3.2.2 Payoffs and incentives structure

In order to ensure equal effort across both the sections, instructions explicitly stated that each section had a 50% chance of being picked up for the payment. It was ensured that students understood payoff scheme as given in Panel B, Table 1 before starting the quiz. Participation honorarium was contingent on completing the task inclusive of post-quiz questionnaire. Students were informed about the payoff structure during the instructions. Over and above a participation honorarium of INR 150, students could earn 0 (getting all 20 questions wrong) to 400(getting all 20 questions right) depending upon their performance in the randomly selected section post task. Subjects in treatment (control) session were told that they would be awarded 20 (20) points for each correct answer, -10 (0) for each incorrect answer,and 0 (0) points for skipping. The payoffs from the quiz was capped at 0 from below

[19]To elucidate further, Matrix and BODMAS had three high effort types and two low effort types questions; Linear equation and Basic Probability had three low effort types and two high effort types. Effort type is in terms of thinking/time/steps required to solve a question. For example of the two types see Appendix E.

[20]Students typically undertake Graduate Record Examinations (GRE) for applying to foreign universities after post-graduate level and National Eligibility Test(NET) is undertaken by students after post-graduation if they are interested in pursuing teaching at graduate level or Ph.D. in Indian Universities. It is unlikely that under graduate students would have exposure to these set of questions beforehand.

for both the sessions. Hence, the key difference in the treatment and control group was the scoring rule for the wrong answers. To make the quiz more competitive and ensuring best effort across both sections, it was announced that the students in the top quartile (i.e., top 25% students with highest number of correct answers (i.e., total number of correct answers out of 40) would get a bonus of INR 100).

Each section had a fixed twelve minutes time slot. Subjects could neither allocate timings across the sections according to their preference or alternate between sections. Instructions before the start of the quiz clearly described the two sections, time limits etc. so that students were not surprised or overwhelmed by the performance in the first section.

3.2.3 Post task questionnaires

Post quiz, students undertook two more sections - Section (iii) Participant's characteristics and Section (iv) Post quiz questionnaire (see Appendix F). While Section (iii) consisted of questions on personal characteristics like gender, religion, CGPA (Cumulative Grade Point Average), etc., and a standard Loss Aversion Task.[21] Section (iv) consisted of questions that aimed at measuring subject's self perception about own and other's performance and confidence levels based on their perceived self-assessment of the quiz performance.

4 Data and results

4.1 Summary Statistics

Panel A, Table 1 gives the sample distribution by treatment type across the two institute. Since Institute 1 is the larger institute, larger proportion of the sample comes from here. Out of 275 total subjects, 53% (i.e., 146) were in sessions with

[21] We use a measure which is a standard in loss aversion literature **???**. Karle et al. (2022) further show that these measures are a good predictor of an individual's actual loss aversion.

penalty treatment. Around 39% (i.e., 107) of our participants identified themselves as females.

Panel B of Table 1 gives details of the expected payoff from the study which was same across the treatments.

Table 2 shows that the subjects are balanced on the observables in the treatment and control groups, the only exception being proportion of subjects with English as medium of instruction at the primary level of schooling, albeit at 10% level of significance. Majority of our subjects were pure B.Tech degree students as opposed to dual degree students, and identified themselves as Hindus, being from general administrative caste category, had English as medium of instruction at primary level of schooling, with a mean age of 19.6 years and had spent about two years in the current institute. Students came from about 26 institute-department combination in varied proportion. In Table 2, we report that 30% of the our participants were from the three most popular(sought after) department by engineering aspirants. Around 61% of students out of total of 275 students reported their CGPA to be above 7 points, 11% chose not to report their CGPA. Nevertheless, we conduct our analysis using variables reported in Table 2 as controls.

Next, we define our outcomes of interest viz. performance in the quiz. We proxy performance in the quiz by looking at (i) net scores (20*total number of correct answers - penalty*total number of incorrect answers, where penalty takes value '10' if an individual 'i' is in penalty treatment and '0' if in non-penalty treatment; (ii) Ranks based on net scores, a higher rank implies better scores (like CGPA), (iii) Number of attempts (total number of question attempted), (iv) Number of correct answers, (v) Unconditional accuracy rates (number of correct answers out of given questions); (vi) Conditional accuracy rates (number of correct answers out of number of questions attempted). We have summarized outcomes in Table 3.

As seen from the column (2) and column (3), students performance is significantly lower under penalty treatment when we look at net scores and ranks. Main impact comes from number of attempts – a significant difference of 5 questions across the two treatment arms. Number of correct answers reduce by two answers in penalty treatment, results being driven by Logical Reasoning(LR) section. Interestingly, there is no mean difference in number of correct answers in the Simple Arithmetic (SA) section that had familiar and technically lesser advanced questions. We observe a fall in unconditional accuracy rates but a significant increase in conditional accuracy rates, i.e., precision rates seem to be improving with the introduction of the penalties.

Figure 1 breaks Table 3 by treatment and gender. Mean and confidence interval plots show that while within treatment differences for two genders are small and significant, females in penalty are most affected especially when we look at number of attempts and have highest conditional accuracy rates. Notably, women in non-penalty are performing best and have highest unconditional accuracy rates (albeit insignificant as compared to men in non penalty treatment). Penalty affects both males and females but affect females more.[22]

Below we describe the methodology and the estimating equation that will give us the causal impact of penalty treatment on gender.

4.2 Empirical Methodology

We use OLS (Ordinary Least Squares) regression analysis to causally estimate the impact of negative marking on the participant's outcomes. Our baseline specification is as follows.

[22] **??** gives break up of outcomes by gender. Females have higher total scores, albeit insignificant plausibly because two sections have opposite results. Females perform better in LR because they have higher unconditional and conditional accuracy rates that pulls up their scores despite attempting similar number of questions as men.

$$\mathrm{Y}_i = \alpha + \beta_1 Treatment_i + \beta_2 Gender_i + \beta_3 Treatment_i * Gender_i + \gamma \mathrm{X}_i + \epsilon_i \quad (1)$$

, where Y_i are the outcomes of interest as defined in Table 3. β_1 and β_2 gives differential impact of scoring rule ($Treatment_i$ takes value 1 if individual *'i'* was in the session with penalty treatment, 0 otherwise) and gender ($Gender_i$ takes value 1 if an individual identified themselves as female) on outcomes, respectively. β_3 is our main coefficient of interest and gives us causal impact of negative marking on females' performance.

Table 2 renders confidence to our estimation strategy as it shows balancing on observables across treatment and control groups at 5% level of significance. Nevertheless we control for observables as described in Table 2.[23] Additionally, our post task questionnaire also collected information on participant's degree of loss aversion, self perception, confidence and beliefs about own and other's performance that have been shown in differ systematically by gender and may have bearing in a MCQ-negative marking based setting. We conduct analysis with these variables to pin down the mechanism through which negative marking may manifest itself.

All possible measures were taken to inhibit cheating, communication and performance spillovers within and across the sessions. We conducted multiple sessions back to back on one day in each institute. The likelihood of spillover across institutes are practically non-existence as institutes are located in two different Indian states with more than 400 km apart and our event in each institute was timed in two different semesters.[24]

[23]Significant difference exists in proportion of students whose mode of instruction till primary level was not English across treatment arms at 10% level of significance. We check for results controlling for primary medium of education and other personal characteristics in our analysis.

[24]We show empirically that there were no spillovers across the two institutes by conducting institute-wise

Our baseline specification also includes covariates described in Table 2. We augment Equation 1 by including institute-branch fixed effects and cluster standard errors at the session level.[25] Results are described in the following sections.

4.3 Results

Results from estimating vanilla version of our baseline specification are given in column (1) to (4) in Table 4. Column (5) to (8) give results for stricter specification that includes characteristics controls (as described in Table 2) and also includes institute-branch fixed effects. Since, ranks under penalty are more likely to be at the lower end of rank distribution by design, we look a them separately in col (3) & (4); (7) & (8). Table 4 shows negative impact of penalty on scores and overall scores. The coefficients against $Female$ are insignificant in all columns but negative and significant against $Penalty * Female$ in column (5) and (6). Females in Penalty end up with lower scores and lower ranks.

We look at other outcomes in Table 5 that lead to outcomes studied in Table 4. Our stricter specification shows stronger results (see col(5) to (8)). Overall, penalty treatment has significant negative impact on number of attempts and positive impact on conditional accuracy rates. There are no overall significant gender differences, but females in penalty attempt about 4 lesser questions and get about 3 more questions incorrect – around 10% fall in attempts and 6% fall in correct questions. However, there is no impact on conditional accuracy rates (precision) even though their unconditional accuracy rates fall.

We break down results of Table 4 and Table 5 by the content tested in the quiz (by the sections). Table 6 shows outcomes in Simple Arithmetic and Logical Reasoning Sections in Panel A and B, respectively. We find strong heterogeneity in the impact

heterogeneity analysis. Details in the section 4.

[25] Since our treatment was at individual-session level, we cannot do session fixed effects or individual fixed effects.

of penalty on the two sections with participants performing worse in the Logical reasoning under penalty treatment (see coefficient against *Penalty* in Panel A and B). There are no gender differences in the performance in either of the sections (see coefficient against $Female$ in Panel A and B). Our key coefficient $Penalty*Female$ is negative and significant for Simple Arithmetic Section but not for Logical Reasoning section. In fact, $Penalty*Female$ is positive for logical reasoning, albeit insignificant.

Focusing on Panel A, we see that Females have significantly lower ranks under *Penalty* (see col (4), Panel A) but no differences on any other outcomes in the Simple Arithmetic section. Intriguingly, the differential impact of penalty on gender ($Penalty * Female$) is entirely driven by the Arithmetic Section. Females in penalty treatment attempt around 4 lesser questions (18%) and get around 3 more questions wrong (15%) as compared to everyone else whereas no such differences exist for the qualitative section.

4.4 Robustness and Heterogeneity checks

4.4.1 Dropping missing values

Overall, we covered 275 B.Tech students across various department in the two institutes. 11% of students preferred not reporting their CGPAs. Running a simple Probit model to check whether these students differ significantly from others (those with reporting their cgpa), we find chi-square for overall fit to be significant and thus, we don't drop these students from our main analysis. However, we cannot use control for academic performance (CGPA levels) for full sample. In addition to this, around 9.8% students answered loss aversion lottery question in an irrational manner. Thus, our sample with full information falls to be 221 students. Table A.2 shows coefficients for all controls for the sample with information on all variables. It is quite interesting to observe that females get an extra half a question correct as compared to everyone

else in Simple Arithmetic section (see column 6 in Table A.2. Apart from this, results are qualitatively and quantitatively similar to our main results in Table 4 and Table 5.[26]

4.4.2 Question fixed effects

Next, we transform data at the individual-question level and estimate a probit model by modifying Equation 1 where dependent variable is a dummy taking value 1 if a question was attempted (or answered), 0 otherwise. This specification gives us probability of attempting a question and getting a question correct as shown in (Table A.4). Column (4) to (6) shows estimates for the strictest specification, viz. using question fixed effects. Again, our main results are upheld. Females in penalty are lesser likely to attempt a question and lesser likely to get a question correct in Simple Arithmetic section (Panel A) with no such significant differences in Logical Reasoning section (Panel B).

4.4.3 Heterogeneity by Effort type

We show results by heterogeneity in estimated effort by our task's design. As mentioned earlier, quiz had questions requiring different level of effort in terms of steps and time required to read and solve the questions. We find consistent impact of penalty for females across the effort type for both the sections in line with the main results. Table A.4 shows that $Penalty * Female$ has significantly lower probability of attempting and getting question correct out of total with no significant impact on getting correct answer conditional on attempting irrespective of question being low or high effort type. Further, no significant impact on probabilities in LR section by effort type as shown by Panel B, Table A.4, although penalty has larger impact on high effort type in LR section (notice that coefficients against *Penalty* increase

[26]Results are similar for the subset of 245 students (individuals with CGPA and other information). For brevity, we have presented result for the strictest sample. Results available on request.

considerably in col (4) and col (5) as against col (1) and (2), respectively. All results hold with question fixed effect.[27]

4.4.4 Heterogeneity due to fatigue or time pressure

We also check if there is any heterogeneity in the impact of penalty due to fatigue or time pressure. We conduct sub sample analysis of first ten questions and last ten questions per section and find no heterogeneity due to the order of questions. Results given in Table A.5.

4.4.5 Heterogeneity by Institute

As mentioned earlier, both institutes are high stake STEM institute in our sample, admitting students through common entrance exam and counseling, there is still a considerable gap between the ranking of the two institutes with students with relatively better ranks on an average going to Institute 1. We break sample by institute and show results in Table A.6. While our main section-wise results are upheld across both institutes, coefficients of interest are larger in magnitude for Institute 2. In fact, if we look at simple arithmetic section, in institute 2, females in penalty treatment attempt 5 less questions as opposed to 2 questions in institute 1 (see col 5). Clearly, impact of negative marking is more pronounced for females in lower rank institute, especially for simple arithmetic section. Notice that conditional accuracy rate is significantly higher for the same (Panel B, col 8, Table A.6).[28]

We summarize our key findings below that are followed by investigation on potential mechanism in the next section:

[27] We also check for validity of our definition of effort type using per question performance. We find that around 66% (76%) of high effort type questions and 80% (82%) of low effort type in SA (LR) were attempted by the subjects. Notably, around 54% (32%) of high effort type and 70% (60%) of low effort types were answered correctly in SA (LR) by the participants.

[28] Stronger results for institute 2 discounts spillover across institutes because we conducted study at institute 2 at a later date.

1. Penalty treatment has a negative impact on overall net scores, number of attempts, number of correct answers but weak impact on strike rates (i.e., accuracy rates).

2. Effects on Penalty are driven by Logical Reasoning section.

3. Females in Penalty treatment attempt significantly lower number of attempts, lesser correct answers but no impact on their strike rates. This translates into lower net scores and ranks for them.

4. Effects on Females in Penalty treatment are driven by their performance in Simple Arithmetic Section.

5 Investigating potential mechanisms

The design of our experiment kept examination method (MCQ objective question type), mode of examination (computer based), preparation level (subjects were unaware of the topics beforehand) and time pressure uniform across subjects and overcame self selection of females into penalty based competitive examinations to give us the causal impact of negative marking based scoring rule. Our data analysis shows that negative marking - a key feature of competitive exam, has potential to generate gendered outcomes. Females in penalty attempt 4 questions less, get around 3 less questions correct but with no significant improvement in conditional accuracy rates, and this translates into lower scores and rank for them. A fall in correct answers by approximately 7% and in attempts by 10% is quite considerable given the context of highly competitive exams where even a point difference has the potential to alter one's entire career trajectory. Most intriguing aspect of our result is heterogeneity by the subject matter. Estimates shoot up to approximately 15% to 20% when we look at Simple Arithmetic section which was supposedly familiar and null effects in

the logical reasoning section which was unfamiliar and technically more advanced by design.

While the impact of penalty is clear, the reasons for the impact and heterogeneity in impact by the content is not straightforward. In the remaining section we discuss some of the plausible mechanisms that may get manifested with negative marking and explain why women skip more in penalty treatment (i.e, attempt less).

5.1 Difference in ability or subject knowledge by gender

Our randomization helped in achieving balance as seen from Table 2 and results are upheld with institute-branch fixed effects and CGPA which are a proxy for general ability or effort or motivation.[29] The only way differences in abilities would explain our results if low ability women got systematically assigned to the penalty treatment and also low ability women are equally good at LR. While randomization is supposed to take care of it, it would also be stretching too far because LR was a more rigorous section and females in penalty treatment have highest mean strike rates (see Figure 1). Further, strike rates for women in penalty are not significantly different in the regression framework (see col 8, Table 5 to Table 8) discounting difference in knowledge.

5.2 Differences in Loss Averseness

We use concept of Loss aversion given by Prospect Theory (Kahneman Tvresky (1992)). We use a standard lottery question that helps us identify how individual value expected losses and gains in a mixed domain with reference point as zero. Lottery question gives us degree of loss aversion called *"point of inflexion"* (Fehr and

[29] Assignment to branches is based on performance in the entrance exam and one's preference may not be true reflection of one's ability – the larger debate that our paper speaks to. It is true that if we were to use entrance exam rank distribution within institute branch level to proxy ability then males have higher ability on average. However, study by Sharma and Sengupta (2024) using academic performance data from the same institutes show that women exit with higher class ranks, opposite of their entry ranks and thus, we do not use entry ranks as a proxy for ability.

Goette (2007), Gachter et al. (2021), and Karle et al. (2015, 2022)).[30] Using when a participant decides to switch from *"I want to toss the coin"* to *"I don't want to toss the coin"* in a lottery with possible gain and loss, we calculate point of inflexion which is equal to the ratio of expected gains to expected loss at the switch. Higher point of inflexion indicates higher loss aversion, i.e., participant values potential loss more than potential gain and thus doesn't want to toss coin thereafter. Participants who never switch or switch multiple times get dropped from the sample because theory of point of inflexion cannot predict degree of loss aversion for them, i.e., their behaviour is irrational in the studied context. We lose about 10% of observations due to this.[31]

Modal point of inflexion in our sample is 1.6, where around 41% participants choose not to toss the coin as they could lose INR 150 or gain INR 250 with a probability of $\frac{1}{2}$. t-test of means shows no significant difference between the degree of loss aversion among participants by treatment or gender.[32] We augment baseline specification to include inflexion point and respective interaction terms and show results in Table 8. Our coefficient of interest is given by $Penalty * Female * InflexionPoint$ and is negative albeit insignificant for SA (see col (1) to (4), Table 8. The results for LR is completely opposite. Col (5) to (8) show females with higher inflexion point get about 3 more questions correct in penalty treatment and have significantly higher unconditional accuracy rates while no differences in attempts. Interestingly, coefficient against $Female * InflexionPoint$ is negative albeit insignificant for SA. While coefficient against $Female * InflexionPoint$ is in line with literature i.e., females with higher degree of loss aversion are attempts less and get fewer questions correct, results with $Penalty * Female * InflexionPoint$ are not straight forward. Unexpect-

[30] See Q8, Appendix F.

[31] Note that 11% did not report their CGPA, including that we lose about 20% of our sample. For completeness sake, we conduct loss aversion analysis on our full information sample.

[32] We use measure of loss aversion instead of risk aversion by Baldiga (2014) because recent papers have critiqued Baldiga (2014)'s measure of risk aversion as nothing but loss aversion measure (Karle et al. (2022)).

edly, coefficients against $Penalty * Female$ for LR become negative and significant as opposite of our main results. To conclude, Loss aversion doesn't entirely explain the difference in performance of females due to negative marking (see coefficient against $Penalty * Female$ for column (5) to (8)).

5.3 Perceived comparative advantage across two sections

Post quiz questionnaire had asked participants to list down the section they would prefer to be considered for payoff calculations. Assuming that a participant would like to maximize payoff, we expect participant to report the subject they feel they have performed better or feel have comparative advantage in the given context. Around 23% participants responded not preferring either section, 59% preferred SA and 17% preferred LR for payment. Breaking these numbers by gender, we see that around 63% of males preferred SA and 14% preferred LR as opposed 53% and 23% for females, remaining being indifferent between the two sections. We carry out sub-sample analysis by preferred section in Table 10 to find that results on SA are driven by females in penalty who are indifferent between the two sections for payment (see col (1), (2) in Panel A, Table 10). For the sub-sample being indifferent between the two sections, the magnitude of impact for $Penalty * Female$ is very high – a fall of 7 (35%) and 5 (25%) in total number of attempts and correct answers. The females who don't report a preference for a section are performing significantly worse than those who respond the same and for females specifying a section (either SA or LR, col (4) to (6) or col (7) to (9)), penalty has no differential impact as compared to others specifying that particular section. Again, we find no heterogeneity in results for LR (see Panel B). These results suggest that female's performance get affected more with penalty when they have ambiguous beliefs or lower confidence.

5.4 Self evaluation across two sections

We measure over estimation or performance gap as the difference between the number of questions participants thought they got right in the respective sections and the actual number of correct answers. We divide the sample between those who overestimated and those who did not.[33] While $Penalty * Female$ is negative and significant in SA section irrespective of whether females overestimate or don't, we see an interesting result. Females in penalty who overestimate have significantly higher conditional accuracy rates (6.6 p.p higher, $p < 0.01$, see col (3), Table 11) and lower number of correct answers but weakly significant (≈ 3, $p < 0.1$, see col (1), Table 11) suggesting that females need higher confidence to close the gap in scores (through total number of correct answers) when faced with penalty for incorrect answers. Results from column (7) to (8) suggest that females shy away from attempting when faced with penalty if they feel less confident but this holds only for SA.

A related variable gauging confidence of the participants is where they place themselves in the rank distribution of getting total number of questions correct among those taking quiz within their institute (see question 3, Section IV, Appendix F). Males rank themselves in better deciles as compared to females with significant t-test differences by gender in the penalty treatment. Further, a significantly higher proportion of males feel that would win a bonus of INR 100 because they would be in top 25%.[34]

To summarize, variables from post-quiz questionnaire *suggest* that penalty has largest bearing on females who are low in confidence or have lower self-beliefs in competitive situations. Intriguingly, it affects performance only in SA.

[33] Around 49.45% of participants overestimate their performance in SA and 70% in LR. 51% (47%) males (females) overestimate their performance in SA and 71% (67%) overestimate their performance in LR. There are no t-test differences in self estimation by gender.

[34] Response to question 4, Section IV, Appendix F.

While we cannot directly test why penalty affects confidence or performance only in SA, we turn to the existing stereotype/maths anxiety literature that shows that women's performance in maths gets adversely affected when they know that they are going to evaluated on the basis of their performance in mathematics. Research has used this evidence to explain gender disparity in STEM courses uptake as well (Helene et al (2023), K. Morrissey, et al (2019), Hyde et al. (2008), Hyde & Mertz (2009), Farenga & Joyce (1999), Eccles & Jacobs, (1986); Fennema & Sherman (1978)).[35]

Another explanation could be that MCQ tests short term working memory as suggested by (Singh and Chakravarty (2018)). Since in our context, SA particularly requires recalling formula and needed short term working memory (at which men might also have advantage due to coaching culture) as compared to logical reasoning, we see significant negative effects with penalty for females.

Irrespective of the mechanism, results are surprising because we tested technically easier and familiar questions in SA and our sample consisted of females who have self selected to purse Engineering for graduation and have cleared one of the toughest quantitative exam. Thus, our estimates are arguably lower bounds in performance gap due to penalty and would intensify further with other inequalities.

6 Conclusion

We conducted a controlled experiment that randomly assigned participants to negative marking based scoring rule and no negative marking based scoring rule in a mixed effort quiz consisting of two dissimilar section. We find that negative marking impacts females deferentially by reducing their number of attempts and number of correct answers. However, their precision remains same. This results in lower net scores and lower ranks having direct implications in competitive exams that use

[35]This is also suggestive of difference in take up of NEET which doesn't test maths inducing girls to aspire to be a medical professional.

negative marking-MCQ for screening candidates. Results are driven by significant differences in female's performance in the familiar and technically less advanced Simple Arithmetic Section as opposed to the logical reasoning section when faced with penalty. We argue that stereotypes (maths anxiety) explain heterogeneity by the subject matter best. Penalty induces females to attempt in a generally perceived to be male domain even though it was technically less advanced. Given the context of competitive exam where a difference of even a quarter of a point matters, it is important to be cognizant of neutral subject matter and scoring scheme for inclusivity while dsigning the assessment process. Our results also advocate early childhood un-conditioning of maths as a male domain. Moreover, future studies should focus on designing and testing gender neutral assessment schemes, for example confidence assessment.

References

Baldiga, K. (2014): "Gender Differences in Willingness to Guess," *Management Science*, 60, 434–448.

Bliss, L. B. (1980): "A test of Lord's assumption regarding examinee guessing behavior on multiple-choice tests using elementary school students," *Journal of Educational Measurement*, 17, 147–154.

Burns, J., S. Halliday, and M. Keswell (2012): "Gender and Risk Taking in the Classroom," *Southern Africa Labour and Development Research Unit*, Working Paper Series No.87.

Coffman, K. B. and D. Klinowski (2020): "The impact of penalties for wrong answers on the gender gap in test scores," *Proceedings of the National Academy of Sciences*, 117, 8794–8803.

Espinosa, M. P. and J. Gardeazabal (2005): "Do students behave rationally in multiple-choice tests? Evidence from a field experiment," *Working paper*.

——— (2010): "Optimal correction for guessing in multiple-choice tests," *Journal of Mathematical Psychology*, 54, 415–425.

Graetz, G. and A. Karimi (2022): "Gender gap variation across assessment types: Explanations and implications," *Economics of Education Review*, 91.

Jain, T., A. Mukhopadhyay, N. Prakash, and R. Rakesh (2021): "Science education and labor market outcomes in a developing economy," *Economic Inquiry*.

Karle, h., D. Engelmann, and M. Peitz (2022): "Student performance and loss aversion," *Scandinavian Journal of Economics*.

Riener, G. and Wagner (2017): "Shying away from demanding tasks? Experimental evidence on gender differences in answering multiple-choice questions," *Economics of Education Review*.

Saygin, P. and A. Atwater (2021): "Gender differences in leaving questions blank on high-stakes standardized tests," *Economics of Education Review*, 84.

Şahoo, S. and S. Klasen (2018): "Gender Segregation in Education and its Implications for Labour Market Outcomes: Evidence from India," *IZA Discussion Paper No. 11660*.

Table 1: Sample sizes across Gender and Treatment

Panel A: Sample Distribution			
	No Penalty	Penalty	Total
Female			
I1	27	37	64
I2	22	21	43
Total	**49**	**58**	**107**
Male			
I1	42	44	86
I2	38	44	82
Total	**80**	**88**	**168**
Total	**129**	**146**	**275**

Panel B: Payoff Scheme in INR			
Participation honorarium	Performance based	Bonus lottery	**Final amount**
150	{0,400}	{0,100}	**{150,650}**

Notes:INR=Indian Rupees. 275 participants were spread across seven sessions of different size

Table 2: Participant's characteristics by treatment arms

Characteristics	Penalty	No-Penalty	Difference
(1)	(2)	(3)	(4)
N=275	146	129	(3)-(2)
Prop. of women	0.397	0.380	-0.017
	(0.041)	(0.043)	(0.059)
Age	19.565	19.783	0.218
	(0.100)	(0.143)	(0.171)
Prop. of Hindu	0.822	0.883	-0.061
	(0.032)	(0.029)	(0.043)
Prop. reported general caste	0.473	0.539	0.066
	(0.041)	(0.044)	(0.061)
Prop. with English medium primary schooling	0.904	0.961	0.057*
	(0.024)	(0.017)	(0.031)
Prop. in B.Tech. program	0.938	0.899	-0.039
	(0.020)	(0.027)	(0.033)
Prop in popular disciplines@	0.301	0.380	0.078
	(0.038)	(0.043)	(0.057)
No. of years in institute	1.849	2.008	0.158
	(0.091)	(0.117)	(0.146)
Prop. CGPA>7#	0.664	0.718	0.054
	(0.042)	(0.042)	(0.059)

Notes @Selective disciplines in B.Tech, IITs=Computer Science & engineering + Electrical Engineering+Mathematics and Computing (source: Sharma & Sengupta 2024), #For N=245, 11% preferred not to report their CGPAs. Standard errors in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Outcomes of interest by Treatment Arms

Outcomes	Overall	Penalty	No-Penalty	Difference
N	(1) 275	(2) 146	(3) 129	(4) (3)-(2)
Net Scores:				
Total	395.818	343.767	454.729	110.962***
SA	240.582	227.534	255.349	27.815***
LR	155.236	116.233	199.380	83.147***
Ranks:				
Total	138	106	174	68***
SA	138	123	154	30***
LR	138	96	185	89***
Attempts:				
Total	30.655	28.274	33.349	5.075***
SA	14.756	14.151	15.442	1.291***
LR	15.898	14.123	17.907	3.784***
Correct:				
Total	21.753	20.884	22.736	1.853***
SA	12.520	12.301	12.767	0.466
LR	9.233	8.582	9.969	1.387***
Unconditional Accuracy rates:				
Total	0.544	0.522	0.568	0.046***
SA	0.626	0.615	0.638	0.023
LR	0.462	0.429	0.498	0.069***
Conditional Accuracy rates:				
Total	0.710	0.738	0.679	-0.059***
SA	0.852	0.874	0.827	-0.047***
LR	0.580	0.600	0.558	-0.042**

Notes: Attempts= count of number of question attempted; Correct= number of correct answers; Unconditional Accuracy rates= Number of correct answers out of assigned questions; Conditional Accuracy rates= Number of correct answers out of attempted questions; Gross Scores= 20*number of correct answers; Net Scores=20*number of correct answers -penalty*number of incorrect answers; Ranks=rank(number of correct answers) where higher score maps into higher rank. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Net scores and Ranks

		Ranks				Ranks		
	Net Scores	(Overall)	(No Penalty)	(Penalty)	Net Scores	(Overall)	(No Penalty)	(Penalty)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Penalty=1	-98.955***	-60.560***			-90.216***	-55.751***		
	(24.334)	(14.004)			(9.452)	(5.941)		
Female=1	16.398	9.605	3.176	-7.108	4.395	3.254	1.602	-12.552
	(12.145)	(6.895)	(4.972)	(5.759)	(5.465)	(3.143)	(3.233)	(9.140)
Penalty*Female	-30.943	-19.548			-40.558**	-24.932*		
	(17.299)	(11.728)			(15.023)	(11.687)		
Characteristics controls					✓	✓	✓	✓
Institute-Branch FE.					✓	✓	✓	✓
Constant	448.500***	170.537***	63.794**	76.324***	548.817***	224.538**	44.721	176.163*
	(17.813)	(11.400)	(6.724)	(5.298)	(140.198)	(85.187)	(18.393)	(68.285)
N	275	275	129	146	274	274	128	146
R-sq	0.187	0.187	0.002	0.007	0.467	0.462	0.417	0.431

Net Scores=20*number of correct answers -penalty per question*number of incorrect answers where penalty=0 for no penalty treatment, =10 for penalty treatment; Ranks=rank(number of correct answers) where higher score maps into higher rank. Characteristics controls include age, dummy for general caste, Hindu religion, English medium primary and number of years in the institute. Robust standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Other outcomes

	Total		Accuracy rates		Total		Accuracy rates	
	Attempts	Correct	Unconditional	Conditional	Attempts	Correct	Unconditional	Conditional
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Penalty=1	-3.809**	-0.993	-0.025	0.059***	-3.273***	-0.479	-0.012	0.063***
	(1.040)	(1.166)	(0.029)	(0.015)	(0.681)	(0.429)	(0.011)	(0.014)
Female=1	0.523	0.820	0.020	0.018**	0.229	0.279	0.007	0.008
	(0.753)	(0.607)	(0.015)	(0.005)	(0.598)	(0.296)	(0.007)	(0.005)
Penalty*Female	-3.209**	-2.200*	-0.055*	-0.001	-3.823***	-2.754**	-0.069**	-0.008
	(1.161)	(0.923)	(0.023)	(0.009)	(1.026)	(0.751)	(0.019)	(0.015)
Characteristics controls					✓	✓	✓	✓
Institute-Branch FE.					✓	✓	✓	✓
Constant	33.150***	22.425***	0.561***	0.672***	41.877***	27.453***	0.686***	0.654**
	(0.794)	(0.891)	(0.022)	(0.012)	(10.182)	(6.480)	(0.162)	(0.194)
N	275	275	275	275	274	274	274	274
R-sq	0.175	0.036	0.036	0.073	0.407	0.386	0.386	0.255

Attempts= count of number of question attempted; Correct= number of correct answers; Unconditional Accuracy rates= Number of correct answers out of assigned questions; Conditional Accuracy rates= Number of correct answers out of attempted questions. Standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Outcomes by Section type

		Ranks					Accuracy rates	
	Net Scores	(Overall)	(No Penalty)	(Penalty)	Attempts	Correct	(Unconditional)	(Conditional)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Section I - Simple Arithmetic								
Penalty=1	-5.296	-7.627			0.206	0.770*	0.038*	0.039*
	(7.845)	(8.992)			(0.447)	(0.328)	(0.016)	(0.018)
Female=1	-1.189	0.627	2.065	-28.526*	0.185	-0.011	-0.001	-0.013
	(5.146)	(5.223)	(0.799)	(10.280)	(0.535)	(0.231)	(0.012)	(0.017)
Penalty*Female	-51.393***	-55.526***			-3.652***	-2.903***	-0.145***	0.026
	(9.925)	(11.716)			(0.822)	(0.494)	(0.025)	(0.027)
Constant	306.224**	194.654	12.704	204.122*	19.657**	15.586**	0.779**	0.795**
	(124.780)	(120.988)	(9.657)	(76.095)	(6.385)	(5.857)	(0.293)	(0.232)
R-sq	0.372	0.369	0.411	0.457	0.387	0.379	0.379	0.173
Panel B: Section II - Logical Reasoning								
Penalty=1	-84.920***	-88.781***			-3.478***	-1.249**	-0.062**	0.035
	(8.961)	(8.149)			(0.562)	(0.416)	(0.021)	(0.020)
Female=1	5.584	7.113	3.574	8.644	0.044	0.290	0.014	0.012
	(7.723)	(7.100)	(6.110)	(8.748)	(0.775)	(0.429)	(0.021)	(0.006)
Penalty*Female	10.835	7.818			-0.171	0.149	0.007	0.026
	(13.489)	(13.072)			(0.893)	(0.655)	(0.033)	(0.036)
Constant	242.593***	211.606***	79.906	95.060*	22.220***	11.867***	0.593***	0.541**
	(55.630)	(52.622)	(57.105)	(38.859)	(5.724)	(2.866)	(0.143)	(0.189)
R-sq	0.490	0.509	0.356	0.323	0.380	0.322	0.322	0.207
Characteristics controls	✓	✓	✓	✓	✓	✓	✓	✓
Institute-Branch FE.	✓	✓	✓	✓	✓	✓	✓	✓
N	274	274	128	146	274	274	274	274

Note: As elucidated above. Standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Outcomes by Section type

	Attempts	Correct	Accuracy rates	
			Uncondi.	Condi.
	(1)	(2)	(3)	(4)
Panel A: Section I - Simple Arithmetic				
Penalty=1	0.206	0.770*	0.038*	0.039*
	(0.447)	(0.328)	(0.016)	(0.018)
Female=1	0.185	-0.011	-0.001	-0.013
	(0.535)	(0.231)	(0.012)	(0.017)
Penalty*Female	**-3.652*****	**-2.903*****	-0.145***	**0.026**
	(0.822)	**(0.494)**	**(0.025)**	**(0.027)**
Constant	19.657**	15.586**	0.779**	0.795**
	(6.385)	(5.857)	(0.293)	(0.232)
R-sq	0.387	0.379	0.379	0.173
Mean	14.748	12.516	0.625	0.852
Panel B: Section II - Logical Reasoning				
Penalty=1	-3.478***	-1.249**	-0.062**	0.035
	(0.562)	(0.416)	(0.021)	(0.020)
Female=1	0.044	0.290	0.014	0.012
	(0.775)	(0.429)	(0.021)	(0.006)
Penalty*Female	-0.171	0.149	0.007	0.026
	(0.893)	(0.655)	(0.033)	(0.036)
Constant	22.220***	11.867***	0.593***	0.541**
	(5.724)	(2.866)	(0.143)	(0.189)
R-sq	0.380	0.322	0.322	0.207
Mean	15.898	9.241	0.462	0.581
Characteristics controls	✓	✓	✓	✓
Institute-Branch FE.	✓	✓	✓	✓
N	274	274	274	274

Note: As elucidated above. Standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Outcomes with control for Loss Aversion

	Simple Arithmetic				Logical Reasoning			
			Accuracy rates				Accuracy rates	
	Attempts	Correct	Unconditi.	Conditi.	Attempts	Correct	Unconditi.	Conditi.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Penalty=1	-1.804	-0.565	-0.028	0.075	-1.719	1.414	0.071	0.117
	(1.277)	(1.029)	(0.051)	(0.064)	(2.274)	(1.531)	(0.077)	(0.067)
Female=1	0.727	0.562	0.028	-0.020	1.814	3.528**	0.176**	0.146**
	(1.156)	(0.910)	(0.045)	(0.020)	(1.543)	(1.166)	(0.058)	(0.049)
Inflexion Point	-0.616*	-0.591	-0.030	-0.004	0.280	0.376	0.019	0.015
	(0.306)	(0.409)	(0.020)	(0.024)	(0.436)	(0.283)	(0.014)	(0.009)
Penalty*Female	-1.992	-2.185	-0.109	-0.044	-1.642	-4.457*	-0.223*	-0.173*
	(1.956)	(1.834)	(0.092)	(0.067)	(2.495)	(2.064)	(0.103)	(0.075)
Inflexion Point*Penalty	1.330	1.053	0.053	-0.011	-1.073	-1.572	-0.079	-0.047
	(0.814)	(0.865)	(0.043)	(0.033)	(1.386)	(0.867)	(0.043)	(0.028)
Female*Inflexion point	-0.299	-0.222	-0.011	0.013	-1.109	-2.116*	-0.106*	-0.089*
	(1.121)	(0.730)	(0.036)	(0.014)	(1.032)	(0.901)	(0.045)	(0.037)
Female*Penalty*Inflexion Point	-0.964	-0.572	-0.029	0.031	1.096	3.013	0.151	0.126*
	(1.739)	(1.524)	(0.076)	(0.033)	(1.718)	(1.671)	(0.084)	(0.062)
Characteristics controls	✓	✓	✓	✓	✓	✓	✓	✓
Institute-Branch FE.	✓	✓	✓	✓	✓	✓	✓	✓
Constant	21.439**	16.832**	0.842**	0.751**	23.412**	11.980***	0.599***	0.547**
	(6.651)	(6.558)	(0.328)	(0.298)	(6.636)	(2.466)	(0.123)	(0.211)
N	248.000	248.000	248.000	248.000	248.000	248.000	248.000	248.000
R-sq	0.401	0.397	0.397	0.197	0.430	0.368	0.368	0.228

Includes Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Outcomes with control for Loss Aversion and CGPA

	Simple Arithmetic				Logical Reasoning			
			Accuracy rates				Accuracy rates	
	Attempts	Correct	Unconditi.	Conditi.	Attempts	Correct	Unconditi.	Conditi.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Penalty=1	-1.839	-0.061	-0.003	0.113	-2.051	1.056	0.053	0.092
	(1.492)	(1.279)	(0.064)	(0.059)	(2.534)	(1.606)	(0.080)	(0.080)
Female=1	0.991	1.075	0.054	-0.007	1.726	3.350**	0.168**	0.139**
	(2.057)	(2.108)	(0.105)	(0.046)	(1.780)	(1.313)	(0.066)	(0.045)
Inflexion Point	-0.439	-0.245	-0.012	0.009	0.067	0.096	0.005	0.002
	(0.460)	(0.359)	(0.018)	(0.015)	(0.443)	(0.199)	(0.010)	(0.011)
Penalty*Female	-2.317	-2.725	-0.136	-0.061	-1.199	-4.081*	-0.204*	-0.154*
	(2.476)	(2.630)	(0.131)	(0.087)	(2.293)	(1.879)	(0.094)	(0.071)
Inflexion Point*Penalty	1.235	0.698	0.035	-0.030	-0.915	-1.409	-0.070	-0.034
	(0.965)	(1.000)	(0.050)	(0.030)	(1.589)	(0.942)	(0.047)	(0.037)
Female*Inflexion point	-0.341	-0.403	-0.020	0.008	-1.070	-2.196*	-0.110*	-0.093*
	(1.727)	(1.494)	(0.075)	(0.014)	(1.086)	(1.015)	(0.051)	(0.039)
Female*Penalty*Inflexion Point	-0.942	-0.348	-0.017	0.042	0.984	3.053*	0.153*	0.120**
	(2.013)	(1.845)	(0.092)	(0.038)	(1.500)	(1.455)	(0.073)	(0.047)
CGPA	0.691**	1.022***	0.051***	0.026**	0.300	0.509*	0.025*	0.024***
	(0.276)	(0.230)	(0.011)	(0.010)	(0.332)	(0.228)	(0.011)	(0.005)
Characteristics controls	✓	✓	✓	✓	✓	✓	✓	✓
Institute-Branch FE.	✓	✓	✓	✓	✓	✓	✓	✓
Constant	15.602	7.682	0.384	0.455	21.795**	9.762**	0.488**	0.526
	(8.161)	(6.643)	(0.332)	(0.255)	(6.460)	(2.890)	(0.144)	(0.284)
N	221	221	221	221	221	221	221	221
R-sq	0.452	0.482	0.482	0.249	0.408	0.379	0.379	0.247

Includes controls for CGPA. Estimates for sample with full information. Includes Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Section wise outcomes by preferred section for payment

	Prefers neither			Prefers SA			Prefers LR		
	Attempts (1)	Correct (2)	Con. Accuracy (3)	Attempts (4)	Correct (5)	Con. Accuracy (6)	Attempts (7)	Correct (8)	Con. Accuracy (9)
Panel A: Section I - Simple Arithmetic									
Penalty=1	2.024	1.340	-0.027	-1.569*	-0.830	0.033	-0.942	0.336	0.075
	(1.725)	(1.345)	(0.045)	(0.683)	(0.512)	(0.025)	(1.214)	(0.934)	(0.038)
Female=1	1.568	0.270	-0.058	-0.570	-0.393	0.006	-0.700	-0.902	-0.038
	(1.932)	(1.095)	(0.079)	(0.939)	(0.588)	(0.017)	(1.906)	(1.182)	(0.033)
Penalty*Female	-7.107**	-5.049**	0.068	-1.519	-1.432	-0.007	0.164	1.425	0.105
	(2.295)	(1.815)	(0.083)	(1.311)	(1.175)	(0.043)	(3.029)	(1.791)	(0.098)
Constant	27.770**	18.677	0.492	18.913**	15.049**	0.729**	36.047*	39.195**	1.466**
	(8.976)	(11.522)	(1.006)	(5.318)	(4.710)	(0.211)	(15.892)	(14.739)	(0.435)
R-sq	0.645	0.563	0.371	0.425	0.460	0.136	0.629	0.612	0.603
Mean	13.563	11.469	0.858	15.772	13.519	0.860	12.875	10.521	0.819
Panel B: Section II - Logical Reasoning									
Penalty=1	-2.373***	-0.945	0.035	-4.244***	-1.448*	0.050	-1.327*	1.199	0.116
	(0.538)	(0.982)	(0.065)	(0.995)	(0.632)	(0.034)	(0.537)	(1.196)	(0.067)
Female=1	-0.189	-0.715	-0.028	-0.297	-0.067	0.002	0.446	1.175**	0.049
	(0.721)	(0.955)	(0.058)	(0.982)	(0.318)	(0.024)	(1.052)	(0.400)	(0.051)
Penalty*Female	0.444	-0.372	-0.054	0.103	0.954	0.072	0.882	-2.044	-0.167
	(0.598)	(0.983)	(0.068)	(1.537)	(1.123)	(0.059)	(1.824)	(1.879)	(0.099)
Constant	27.600	19.762	0.973	16.850	9.769	0.561	44.245**	16.780*	0.012
	(25.287)	(12.643)	(0.995)	(14.096)	(7.606)	(0.309)	(12.895)	(7.057)	(0.235)
R-sq	0.578	0.553	0.412	0.492	0.371	0.281	0.737	0.801	0.682
Mean	15.203	8.719	0.580	15.734	9.056	0.573	17.375	10.562	0.608
Characteristics controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Institute-Branch FE.	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	64	64	64	162	162	162	48	48	48

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Outcomes by self perception and performance gap (over estimation)

	Did not overestimate own performance						**Overestimated own performance**					
	Simple Arithmetic			Logical reasoning			Simple Arithmetic			Logical Reasoning		
	(1) Attempts	(2) Correct	(3) Cond. Accuracy	(4) Attempts	(5) Correct	(6) Cond. Accuracy	(7) Attempts	(8) Correct	(9) Cond. Accuracy	(10) Attempts	(11) Correct	(12) Cond. Accuracy
Penalty=1	0.910	0.789	0.003	-2.717***	-0.588	0.051	0.506	1.344*	0.053	-3.895***	-1.410	0.046*
	(0.740)	(0.559)	(0.011)	(0.556)	(0.632)	(0.031)	(0.757)	(0.587)	(0.028)	(0.887)	(0.731)	(0.024)
Female=1	0.751	-0.053	-0.036**	-0.400	0.469	0.044*	0.349	0.433	-0.006	0.488	0.008	-0.020
	(1.101)	(1.155)	(0.013)	(0.837)	(0.269)	(0.020)	(0.543)	(0.399)	(0.045)	(0.716)	(0.821)	(0.030)
Penalty*Female	-4.513**	-2.971*	0.066***	-0.368	0.042	0.029	-3.683***	-3.503***	-0.001	0.001	-0.127	-0.006
	(1.330)	(1.265)	(0.014)	(0.953)	(0.843)	(0.055)	(0.864)	(0.838)	(0.071)	(1.127)	(1.375)	(0.060)
Characteristics controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Institute-Branch FE.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Constant	16.951	20.414	1.166***	17.234	15.377	1.111**	22.862	10.618	0.405	29.943*	10.111	0.033
	(11.898)	(10.526)	(0.178)	(10.570)	(8.133)	(0.304)	(12.112)	(8.451)	(0.447)	(14.152)	(8.152)	(0.319)
N	138	138	138	138	138	138	135	135	135	135	135	135
R-sq	0.462	0.432	0.345	0.445	0.391	0.282	0.446	0.448	0.254	0.417	0.370	0.296

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

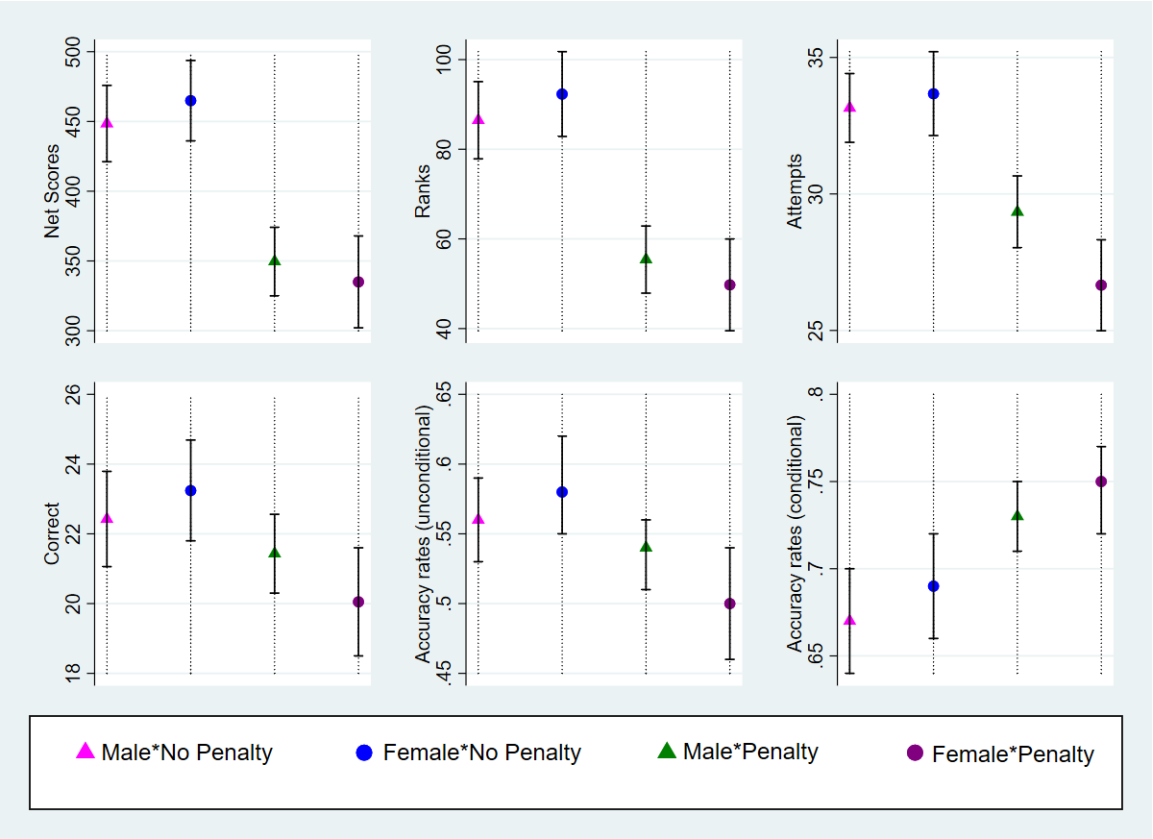
Table 12: Outcomes by belief in ability or luck (imposter syndrome)

	Believes to be a top quartile performer						**Doesn't Believe to be a top quartile performer**					
	Simple Arithmetic			Logical reasoning			Simple Arithmetic			Logical Reasoning		
	(1) Attempts	(2) Correct	(3) Cond. Accuracy	(4) Attempts	(5) Correct	(6) Cond. Accuracy	(7) Attempts	(8) Correct	(9) Cond. Accuracy	(10) Attempts	(11) Correct	(12) Cond. Accuracy
Penalty=1	0.285	0.769***	0.030*	-2.827***	-0.896	0.045**	0.462	1.216	0.078	-6.405***	-2.005	0.061
	(0.269)	(0.201)	(0.012)	(0.597)	(0.461)	(0.016)	(1.853)	(1.959)	(0.047)	(1.147)	(1.179)	(0.093)
Female=1	0.236	-0.218	-0.030	0.260	-0.060	-0.010	-0.735	0.349	0.076**	-1.366	1.507*	0.140
	(0.266)	(0.339)	(0.031)	(0.509)	(0.638)	(0.024)	(2.159)	(2.101)	(0.027)	(2.286)	(0.729)	(0.073)
Penalty*Female	-3.553***	-2.338***	0.067	-0.687	-0.001	0.013	-2.976	-3.910	-0.114	3.190	-0.222	-0.071
	(0.713)	(0.450)	(0.044)	(0.682)	(0.653)	(0.027)	(2.491)	(2.667)	(0.077)	(2.859)	(1.775)	(0.163)
Characteristics controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Institute-Branch FE.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Constant	20.585**	16.134*	0.816*	27.927***	12.912	0.366	27.055**	20.917**	0.694*	25.730**	20.891*	1.290*
	(6.815)	(8.033)	(0.405)	(6.989)	(7.296)	(0.429)	(9.945)	(7.799)	(0.297)	(8.726)	(10.398)	(0.568)
Mean	15.31	12.984	0.852	16.346	9.505	0.538	13.333	11.333	0.851	14.769	8.576	0.574
N	196	196	196	196	196	196	78	78	78	78	78	78
R-sq	0.459	0.425	0.248	0.382	0.324	0.237	0.393	0.432	0.396	0.571	0.576	0.516

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Confidence interval and mean outcomes by treatment*gender



Note: Attempts = Count of attempts out of 40 questions; Correct = Count of correct answers out of 40; Accuracy rates (unconditional) = Count of correct answers out of 40 questions; Accuracy rates (conditional) = Count of correct answers out of total attempts. Net scores = 20*count of correct answers - penalty*count of incorrect answers, where, penalty=0 for non-penalty treatment and 10 for penalty treatment. Ranks = rank(net scores) where higher score means higher (i.e., better) rank.

Appendix A

Table A.1: Outcomes of interest by gender conditional on treatment

Outcomes	**Penalty**			**Non Penalty**		
N	80	49	129	88	58	146
	(1) Male	(2) Female	(3) Diff (1)-(2)	(4) Male	(5) Female	(6) Diff(5)-(4)
Final Scores:						
Total	349.546	335.000	14.545	448.500	464.898	-16.398
SA	242.727	204.483	38.245***	253.750	257.959	-4.209
LR	106.818	130.517	-23.699**	194.750	206.939	-12.189
Ranks:						
Total	109.977	100.035	9.943	170.538	180.143	-9.605
SA	139.926	98.716	41.211**	152.363	157.592	-5.229
LR	87.392	110.112	-22.720*	180.213	192.980	-12.767
Attempts:						
Total	29.341	26.655	2.686***	33.150	33.673	-0.523
SA	15.273	12.448	2.824***	15.325	15.633	-0.308
LR	14.068	14.207	-0.139	17.825	18.041	-0.216
Correct:						
Total	21.432	20.052	1.380	22.425	23.245	-0.820
SA	13.182	10.966	2.216***	12.688	12.898	-0.210
LR	8.250	9.086	-0.836	9.738	10.347	-0.609
Unconditional Accuracy rates:						
Total	0.536	0.501	0.035	0.561	0.581	-0.020
SA	0.659	0.548	0.111***	0.634	0.645	-0.011
LR	0.413	0.454	-0.042	0.487	0.517	-0.030
Conditional Accuracy rates:						
Total	0.731	0.748	-0.017	0.672	0.690	-0.018
SA	0.865	0.889	-0.024	0.829	0.824	0.005
LR	0.578	0.633	-0.056**	0.548	0.575	-0.027

Notes: Attempts= count of number of question attempted; Correct= number of correct answers; Unconditional Accuracy rates= Number of correct answers out of assigned questions; Conditional Accuracy rates= Number of correct answers out of attempted questions; Gross Scores= 20*number of correct answers; Net Scores=20*number of correct answers -penalty*number of incorrect answers; Ranks=rank(number of correct answers) where higher score maps into higher rank. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Outcomes for full information sample

	Overall				Simple Arithmetic				Logical reasoning			
			Accuracy rates				Accuracy rates				Accuracy rates	
	Attempts	Correct	Uncond.	Cond.	Attempts	Correct	Uncond.	Cond.	Attempts	Correct	Uncond.	Cond.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Penalty=1	-3.349***	0.025	0.001	0.077**	0.015	0.995**	0.050**	0.067**	-3.364***	-0.970**	-0.049**	0.045
	(0.662)	(0.677)	(0.017)	(0.023)	(0.441)	(0.341)	(0.017)	(0.019)	(0.442)	(0.384)	(0.019)	(0.031)
Female=1	0.744	0.779	0.019	0.013	0.499	0.500***	0.025***	0.004	0.246	0.279	0.014	0.009
	(0.452)	(0.535)	(0.013)	(0.007)	(0.524)	(0.106)	(0.005)	(0.028)	(0.725)	(0.460)	(0.023)	(0.012)
Penalty*Female	-3.730***	-3.135**	-0.078**	-0.022	-3.799***	-3.314***	-0.166***	0.005	0.069	0.179	0.009	0.013
	(0.757)	(0.955)	(0.024)	(0.027)	(0.621)	(0.305)	(0.015)	(0.035)	(0.690)	(0.675)	(0.034)	(0.050)
Age(yrs.)	-0.696	-0.417	-0.010	0.001	-0.227	-0.142	-0.007	0.005	-0.469	-0.275	-0.014	-0.007
	(0.626)	(0.503)	(0.013)	(0.010)	(0.426)	(0.389)	(0.019)	(0.009)	(0.285)	(0.145)	(0.007)	(0.013)
Reported to be following Hinudism	-0.455	-0.022	-0.001	0.004	-0.514	0.443	0.022	0.053	0.059	-0.465	-0.023	-0.032
	(1.011)	(1.434)	(0.036)	(0.031)	(0.441)	(0.821)	(0.041)	(0.043)	(0.619)	(0.677)	(0.034)	(0.024)
Reported to be from Unreserved caste	2.434**	3.009***	0.075***	0.038**	1.562**	1.653***	0.083***	0.023	0.871**	1.356**	0.068**	0.036
	(0.709)	(0.711)	(0.018)	(0.014)	(0.436)	(0.371)	(0.019)	(0.020)	(0.342)	(0.436)	(0.022)	(0.024)
No. of years in institute	0.965	0.837	0.021	0.007	0.406	0.353	0.018	-0.001	0.559	0.484*	0.024*	0.020
	(0.746)	(0.643)	(0.016)	(0.011)	(0.460)	(0.470)	(0.023)	(0.009)	(0.380)	(0.211)	(0.011)	(0.014)
Primary education english medium	2.935*	3.284**	0.082**	0.037	0.736	1.325	0.066	0.044	2.200**	1.959***	0.098***	0.050*
	(1.291)	(1.015)	(0.025)	(0.029)	(0.902)	(0.784)	(0.039)	(0.038)	(0.596)	(0.363)	(0.018)	(0.025)
CGPA	0.961	1.492***	0.037***	0.026***	0.684*	1.016***	0.051***	0.026*	0.276	0.476*	0.024*	0.022***
	(0.514)	(0.306)	(0.008)	(0.007)	(0.284)	(0.226)	(0.011)	(0.011)	(0.345)	(0.228)	(0.011)	(0.005)
Institute-Branch FE.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Constant	36.405**	16.696	0.417	0.442	14.560	7.033	0.352	0.478*	21.846***	9.662**	0.483**	0.512*
	(12.193)	(8.827)	(0.221)	(0.232)	(8.075)	(6.909)	(0.345)	(0.238)	(5.650)	(2.664)	(0.133)	(0.263)
N	221	221	221	221	221	221	221	221	221	221	221	221
r2	0.467	0.475	0.475	0.320	0.448	0.480	0.480	0.244	0.400	0.355	0.355	0.234

30 students (11%) abstained from reporting CGPA. 27 students (9.8%) answered loss aversion questions inconsistently. 2 students (0.755%) filled incomplete post task questionnaire. Accounting for overlaps in sources of missing information, our sample with complete information becomes 221 from 275. Standard errors in parentheses. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Probability of attempting and answering

	Attempts	Correct (Unconditional)	Correct (Conditional)	Attempts	Correct (Unconditional)	Correct (Conditional)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Section I - Simple Arithmetic						
Penalty=1	0.023	0.104***	0.161**	0.039	0.137***	0.176*
	(0.070)	(0.040)	(0.078)	(0.083)	(0.049)	(0.094)
Female=1	0.041	0.000	-0.057	0.065	0.010	-0.084
	(0.083)	(0.031)	(0.052)	(0.103)	(0.033)	(0.071)
Penalty*Female	-0.571***	-0.394***	0.090	-0.727***	-0.505***	0.067
	(0.112)	(0.058)	(0.102)	(0.143)	(0.073)	(0.124)
Constant	2.133**	1.195	0.705	2.227**	0.802	0.072
	(1.031)	(0.803)	(0.960)	(1.126)	(0.926)	(1.108)
N	5480	5480	4009	5480	5480	3743
Pseudo R-sq	0.069	0.042	0.021	0.242	0.227	0.174
Panel B: Section II - Logical Reasoning						
Penalty=1	-0.668***	-0.160***	0.116***	-0.749***	-0.207***	0.082
	(0.095)	(0.050)	(0.039)	(0.106)	(0.062)	(0.051)
Female=1	0.039	0.036	0.028	0.053	0.052	0.038
	(0.203)	(0.052)	(0.017)	(0.223)	(0.068)	(0.035)
Penalty*Female	-0.061	0.020	0.063	-0.088	0.019	0.068
	(0.219)	(0.079)	(0.048)	(0.241)	(0.102)	(0.063)
Constant	2.450**	0.401	0.122	2.521**	0.102	-0.081
	(1.098)	(0.351)	(0.534)	(1.186)	(0.445)	(0.657)
N	5480	4356	5480	5480	4356	5480
Pseudo R-sq	0.083	0.022	0.015	0.200	0.240	0.242
Characteristics Controls	✓	✓	✓	✓	✓	✓
Institute-Branch FE	✓	✓	✓	✓	✓	✓
Question FE				✓	✓	✓

Dependent variable in Col (1), Col (4) takes value 1 if question was attempted; Col (2) and (5) takes value 1 if question was correctly answered; 0 otherwise; Col (3) and (6) takes value 1 if question was correctly answered conditional on attempting, 0 if incorrect. Standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Probability of attempting and answering with estimated effort type

	Low effort type			High effort type		
	Attempts	Unc. correct	Con. correct	Attempts	Unc. correct	Con. correct
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Section I - Simple Reasoning						
Penalty=1	0.215**	0.238***	0.168*	-0.066	0.059	0.191*
	(0.106)	(0.050)	(0.101)	(0.070)	(0.051)	(0.111)
Female=1	0.218*	0.219***	0.143*	-0.034	-0.155***	-0.294***
	(0.120)	(0.030)	(0.082)	(0.088)	(0.038)	(0.071)
Penalty*Female	-1.104***	-0.808***	-0.213	-0.512***	-0.274***	0.319**
	(0.204)	(0.100)	(0.152)	(0.114)	(0.058)	(0.133)
Constant	1.664	0.300	0.260	2.117**	1.007**	0.340
	(1.787)	(1.621)	(1.641)	(0.846)	(0.504)	(1.253)
N	2700	2740	1920	2740	2740	1813
Pseudo R-sq	0.333	0.267	0.152	0.160	0.174	0.209
Panel B: Section II - Logical Reasoning						
Penalty=1	-0.528***	-0.123*	0.088	-0.952***	-0.303***	0.074
	(0.104)	(0.066)	(0.061)	(0.116)	(0.083)	(0.089)
Female=1	0.148	0.130	0.083	-0.046	-0.023	-0.004
	(0.213)	(0.083)	(0.071)	(0.242)	(0.125)	(0.125)
Penalty*Female	-0.122	-0.003	0.075	-0.040	0.032	0.047
	(0.254)	(0.118)	(0.083)	(0.251)	(0.185)	(0.189)
Constant	2.232	1.143	1.197	2.612***	-0.228	-0.379
	(1.590)	(0.939)	(1.103)	(0.969)	(0.634)	(0.789)
N	2730	2740	2228	2730	2740	2097
Pseudo R-sq	0.219	0.202	0.155	0.189	0.200	0.214
Characteristics Controls	✓	✓	✓	✓	✓	✓
Institute-Branch FE	✓	✓	✓	✓	✓	✓
Question FE	✓	✓	✓	✓	✓	✓

Note: As elucidated above. Low effort type means questions that required one step (lesser time) to solve. High effort type means questions that required multiple steps (more time) to solve. Unc. correct= Probability of being correct, Con. correct=Probability of being correct conditional on attempting. Standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Probability of attempting and answering with first ten (SA)

	First 10 questions			Last 10 questions		
	Attempt	Unc. Correct	Con. Correct	Attempt	Unc. Correct	Con. Correct
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Section I - Simple Arithmetic						
Penalty=1	-0.088**	0.041	0.152	0.105	0.201**	0.220*
	(0.037)	(0.038)	(0.110)	(0.143)	(0.097)	(0.119)
Female=1	0.165***	0.021	-0.144	0.024	0.008	-0.052
	(0.028)	(0.117)	(0.224)	(0.167)	(0.135)	(0.107)
Penalty*Female	-0.660***	-0.335*	0.220	-0.804***	-0.645***	-0.060
	(0.089)	(0.179)	(0.279)	(0.250)	(0.214)	(0.133)
Constant	1.880	0.770	0.649	1.943	0.931	1.038
	(1.167)	(1.204)	(1.923)	(1.646)	(1.151)	(1.368)
N	2740	2740	2024	2740	2730	1711
Pseudo R-sq	0.272	0.252	0.224	0.184	0.141	0.122
Panel B: Section II - Logical Reasoning						
Penalty=1	-1.272***	-0.227***	0.124	-0.451***	-0.188***	0.044
	(0.144)	(0.080)	(0.091)	(0.110)	(0.065)	(0.086)
Female=1	-0.277	0.024	0.079	0.234	0.085	-0.018
	(0.218)	(0.088)	(0.067)	(0.263)	(0.059)	(0.037)
Penalty*Female	0.108	0.005	0.037	-0.145	0.035	0.112
	(0.219)	(0.177)	(0.187)	(0.296)	(0.124)	(0.149)
Constant	2.197**	-0.635*	-0.583	2.821*	1.596**	1.388
	(1.094)	(0.359)	(0.653)	(1.667)	(0.814)	(0.990)
N	2730	2740	2241	2730	2740	2115
Pseudo R-sq	0.272	0.196	0.186	0.178	0.294	0.307
Characteristics Controls	✓	✓	✓	✓	✓	✓
Institute-Branch FE	✓	✓	✓	✓	✓	✓
Question FE	✓	✓	✓	✓	✓	✓

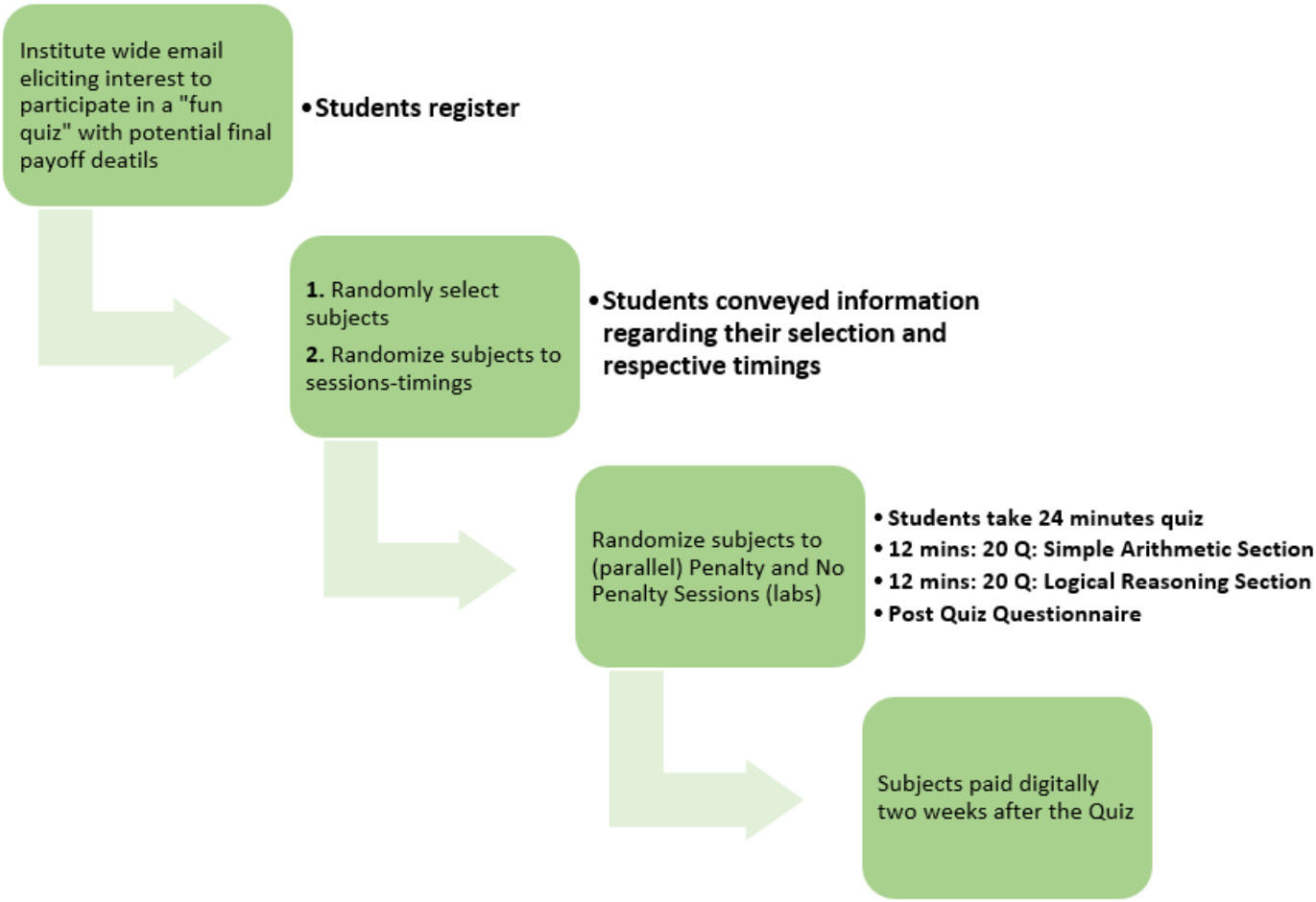
Note: As elucidated above. Standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6:

	Overall				Simple Arithmetic				Logical reasoning			
	Attempts	Correct	Accuracy rates		Attempts	Correct	Accuracy rates		Attempts	Correct	Accuracy rates	
			Unconditional	Conditional			Unconditional	Conditional			Unconditional	Conditional
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Institute 1 (older and relatively higher ranked)												
Penalty=1	-3.819*	-0.829	-0.021	0.058*	-0.557	0.245	0.012	0.043	-3.261**	-1.074	-0.054	0.026
	(1.595)	(0.729)	(0.018)	(0.024)	(0.696)	(0.418)	(0.021)	(0.032)	(1.101)	(0.719)	(0.036)	(0.034)
Female=1	0.259	0.487	0.012	0.011	-0.615	-0.271	-0.014	0.014	0.873	0.758	0.038	0.007
	(1.285)	(0.396)	(0.010)	(0.014)	(0.471)	(0.170)	(0.009)	(0.026)	(0.836)	(0.358)	(0.018)	(0.009)
Penalty*Female	-3.017	-2.291*	-0.057*	-0.000	-2.277*	-2.259**	-0.113**	-0.010	-0.740	-0.032	-0.002	0.060
	(1.917)	(0.837)	(0.021)	(0.025)	(0.873)	(0.592)	(0.030)	(0.045)	(1.321)	(0.973)	(0.049)	(0.050)
Constant	51.526**	29.154**	0.729**	0.558	27.769***	20.088*	1.004*	0.704	23.757	9.066	0.453	0.417
	(14.292)	(9.618)	(0.240)	(0.399)	(5.853)	(7.447)	(0.372)	(0.497)	(11.605)	(5.606)	(0.280)	(0.355)
N	149	149	149	149	149	149	149	149	149	149	149	149
r2	0.428	0.368	0.368	0.320	0.380	0.377	0.377	0.219	0.425	0.347	0.347	0.287
Panel B: Institute 2 (newer and relatively lower ranked)												
Penalty=1	-3.347*	-0.574	-0.014	0.067*	0.637	1.036*	0.052*	0.035	-3.984***	-1.610**	-0.080**	0.042
	(0.439)	(0.095)	(0.002)	(0.006)	(0.391)	(0.164)	(0.008)	(0.010)	(0.048)	(0.069)	(0.003)	(0.008)
Female=1	0.169	0.114	0.003	0.008	0.979**	0.254	0.013	-0.040	-0.810**	-0.139	-0.007	0.017
	(0.073)	(0.059)	(0.001)	(0.002)	(0.031)	(0.132)	(0.007)	(0.010)	(0.042)	(0.073)	(0.004)	(0.004)
Penalty*Female	-4.866**	-3.761**	-0.094**	-0.030**	-5.117***	-3.732***	-0.187***	0.060*	0.251	-0.029	-0.001	-0.028
	(0.281)	(0.224)	(0.006)	(0.002)	(0.025)	(0.058)	(0.003)	(0.009)	(0.305)	(0.166)	(0.008)	(0.004)
Constant	31.467	25.454	0.636	0.756**	12.204	10.956	0.548	0.830**	19.263	14.498	0.725	0.720
	(23.453)	(14.312)	(0.358)	(0.045)	(14.564)	(11.366)	(0.568)	(0.036)	(8.890)	(2.946)	(0.147)	(0.139)
N	125	125	125	125	125	125	125	125	125	125	125	125
r2	0.408	0.401	0.401	0.181	0.371	0.360	0.360	0.162	0.383	0.315	0.315	0.121
Characteristics Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: As elucidated above. Standard errors clustered at session level in parenthesis. Significant at 10% l.o.s * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure A.1: Experiment Flow



Note: Experiment within an institute right from registration to payment to the subjects took about a month.

APPENDIX B: Registration form

This is the registration form for participation in an inter-disciplinary study funded by CoImpact at IITD.

- involves participation in a set of simple non-academic tasks on XX
- you can earn upto INR 650 in approximately one hour
- includes INR 150 participation fess, conditional on completion of the study
- no foreseeable risks
- all your responses, information and performance will be kept confidential and anonymized, as per the standard protocols of any experimental study.
- you are free to withdraw at any point of time of the study.

For any doubts you may contact the undersigned:

Dr. Swati Sharma, Cabin B14, 4th Floor,Block IV, IITD

Prof. Nandana Sengupta,SoPP, Block V, IITD

Prof. Sumitava Mukherjee, HSS, 5th Floor, Main Building, IITD

Kindly note that your participation is contingent on being randomly selected from the registered sample. Filling up this form confirms your interest in the study. We will be using contact details filled by you to share participation and venue details. The study is currently open to only Bachelors' students at XXX. You will be paid within two weeks of the study digitally.

Click "Next" to proceed forward.

1. Name
2. Contact Number
3. Email

4. Entry No
5. Age
6. Enrolled in: B.Tech/Dual degree/Others
7. Department registered with
8. Year of admission
9. Gender: Male/Female/Non-Binary/Prefer not to say
10. Kindly check all that apply:

- I understand that I can earn up to INR 650 (inclusive of the participation honorarium).
- I understand that the participation fees of INR 150 is conditional on me completing the study.
- I have a functioning bank account through which I can accept digital payment for this study.

APPENDIX C: Experiment protocol

Protocol Sheet

1. For volunteers especially at the entrance and corridor area:
    - At the entrance desk (where guard sits): ask for ID and tick against name from the participant-session assignment list (people coming for some other slot need to wait for their slot. **"Hand them Instruction sheet"**
    - Main study room – XXX and YYY
    - We are following the 1-XXX:1-YYY rule by gender, e.g: first girl goes to XXX and second goes to YYY and so on, similarly first boy goes to XXX and second goes to YYY.
    - List it on the room assignment sheet to keep a tab of ratio. Enter it on the Instruction Sheet (if there is time).
    - Aim is to have a similar gender ratio in the rooms. So keep a check on that.

Volunteer at the Lab gate

- No mobile phones allowed. If they are carrying a bag, they can put their phone (on airplane mode) in their bag and deposit it on the shelves. If students don't have bag, then ask them to put phones on silent (no vibrator) or airplane mode. and report to Harshdeep. (He will put a slip of their name and entry no. on the phone and put those in one bag.
- Please check room no. on their instruction sheet and confirm that they are in right room

2. Seat students randomly with a gap of one seat i.e., try to seat participants coming consecutively quite apart (like 1st in 1st row, 2nd in 4th row, etc.). Start with a scattered arrangement.
3. Students in each batch need to be seated within 10-15 minutes. Each two rows facing back to each other, will have one volunteer.
4. Students need to log-in with their ******* ID first.
5. Then, they will put in proxy settings for accessing the internet:
   http://********** (for btech students)
   Mozilla (Windows): Tools → Options → Advanced → Network → Settings → Automatic proxy configuration URL
6. After putting in proxy, they need to open some website like YouTube to check if the internet is working.

7. Meanwhile, row volunteers can start distributing participant information sheets (instruction and consent sheet), plus rough work sheet and a pen (if they need it). Request participants to sign the consent sheet.
8. Ensure everybody has internet access.
9. First session should start, i.e.- ***everybody seated and logged in by*** **2:10 PM. Head volunteer will announce instructions loud and clear at this time:**
   "Before we begin, please listen to the instructions carefully. You can read them from the participant information sheet as well and sign consent.……"
   **Clarify any doubt. It is very important that students understand the payoff function.*
10. **Volunteers to collect participant information and consent sheets. Ensure signatures and correct information. Roll no should be in ********** number format.**
11. **Once everybody has submitted the consent sheets, head experimenter will get a heads up from the row volunteers and then write the LimeSurvey Link on the board.**
    Please announce: " As soon as you write this url, it will take you to a landing page. You need to click on "NEXT" on left bottom corner and the quiz will start right away. Timing will start.
12. After students finish the quiz and post quiz questions (on LimeSurvey) let them be seated. (Students might complete at different times. Students will raise their hands after finishing.) Approach students one by one. Collect rough sheets. Ensure they log out of the system. We need this for the next session. Do not engage in chit chat.
13. **No new participants are allowed in the rooms once the LimeSurvey link has been shared.**

<u>Kindly note:</u>

<u>We cannot allow students to enter after instructions have been announced.</u>

If students arrive 15 mins late (after sharing of link) e.g., around 2:30 PM for 1:45 PM reporting time then request them to wait for the next session.

we inter-change treatment and control rooms after second session

Batch	To-do item	Scheduled time
BATCH 1	Reporting time	1:45
	Student id check + room assignment	2:00
	Last call + instructions	2:10
	Collection of informed consent	2:20
	Quiz starts	2:25
	Quiz ends/Post-quiz starts	2:50
	Post-quiz ends	3:05
	Students leave	3:10
BATCH 2	Reporting time	3:15
	Student id check + room assignment	3:30
	Last call + instructions	3:35
	Collection of informed consent	3:45
	Quiz starts	3:50
	Quiz ends/Post-quiz starts	4:15
	Post-quiz ends	4:30
	Students leave	4:35
BATCH 3	Reporting time	4:45
	Student id check + room assignment	5:00
	Last call + instructions	5:05
	Collection of informed consent	5:15
	Quiz starts	5:20
	Quiz ends/Post-quiz starts	5:45
	Post-quiz ends	6:00
	Students leave	6:05

APPENDIX D: Instructions & Consent Sheet

Participant information sheet

Welcome!
Today, you will be participating in a task that has 2 parts - PART 1: Quiz, PART 2: Post quiz survey

PART 1 INSTRUCTIONS:

- The Quiz will consist of 2 sections: (i) Simple Arithmetic quiz (ii) Logical reasoning quiz
- Section (i) and (ii) have 20 questions each where questions vary in effort level randomly and are organized in no particular order.
- You will get at most fixed 12 minutes per section. Once 12 minutes are over, you will automatically move to the 'NEXT' section.
- Note that the payment you receive will be based on your performance in either one of these randomly selected sections.
- Each correct answer is worth 20 points. There is no penalty for skipping a question. There is a penalty of 10 points per incorrect answer.

PART 2 INSTRUCTIONS:

- Upon completing the quiz you will be presented with a Post quiz survey consisting of 2 sections: (iii) Participant characteristics (iv) Post quiz assessment.
- The total expected time for completing the Post quiz survey is about 10-15 minutes.

TIMING INSTRUCTIONS:

- The total expected time of the entire task is 40 minutes (i.e., Quiz + Post quiz survey). The task link will automatically get disabled after 40 minutes.
- If you finish the quiz sections early, you won't be able to submit before 12 minutes, so make sure you don't restart the quiz. We won't be able to count your performance if you do not complete the entire task in 40 minutes.
- If anyone gets logged out of the system, please login again and continue from where you left.

- For any technical issue, please remain seated and raise your hand. Volunteers will help you.
- Ensure that you log out before leaving.

PAYOFF DETAILS:

- Each participant will earn a participation honorarium of INR 150 on completing the study.
- Each participant will also earn between INR 0-400 based on their performance in a randomly selected section of the quiz. Payoff from the quiz has been floored at 0 for all wrong answers.
- Participants who are in the top quartile of scoring distribution (i.e., total number of correct answers out of 40 questions) for the quiz will earn an additional INR 100 bonus amount.
- The maximum possible payoff is INR 150 (for participation) +INR 400 (from the quiz) + INR 100 (bonus) = INR 650.
- Payments for the quiz will be made digitally by 15th October to the bank details provided.

ADDITIONAL INSTRUCTIONS:

- Full participation in the study implies completing both parts- Quiz and Post quiz survey.
- **Kindly do not try to cheat or prompt answers to anybody. You will be disqualified if found doing so.**
- We will be checking your tab history from the CSC database before making payments. It would be considered cheating if we find that any other tab was opened during the task.
- Similarly, calculators are not allowed.
- **We won't be able to help or clarify any specific quiz questions. Answering is entirely up to your understanding of the questions. If there are any questions till now, then please raise your hand.**
- Please remain seated after finishing your study. Wait for volunteers to collect rough sheets. Kindly do not take any study material with you or discuss answers in the CC Lab premises.

Date_____

Informed Consent

Thank you for participating in this study. You have freedom to withdraw from the study at any stage. If you have any doubts regarding the study's objective then please feel free to contact the undersigned

- Dr. Swati Sharma, Principal Project Scientist, HuSS, Cabin 4B-14, Block IV, IITD
- Prof. Nandana Sengupta, SoPP, Block V, IITD
- Prof. Sumitavah Mukherjee, HuSS, 5th Floor, Main Building, IITD

- I am signing this sheet as a proof of consent for the participation in the study.
- I understand the outline of the project and the freedom to withdraw from the study whenever I wish.
- All the instructions are clear and all my doubts regarding the task have been clarified.
- I understand that I will receive the final payoff for the study digitally by 15th Oct 2023.

Full Name_____ Signature_____

Mob No._____ Entry No_____

APPENDIX E: Effort based task (examples)

Section I: Simple Arithmetic

- There are 20 questions in this section.
- You have **atmost 12 minutes** to answer these questions.
- **This section has a 50% chance of being chosen for payment.**
- Each question is worth 20 points. There is no penalty for not answering (No answer option). There is a penalty of -10 points per incorrect answer.

1.Pick the largest multiple of 17 from Matrix A and the smallest multiple of 19 from Matrix B and calculate the product of the two:

Matrix A		
241	181	153
119	207	187
117	189	217

Matrix B		
171	152	147
134	137	47
131	51	90

Choose one of the following answers

Time remaining 00:11:19

- ○ 31977
- ○ 24824
- ○ 28424
- ○ 11327
- ◉ No answer

2. Two times a number, decreased by 12 equals three times the number, decreased by 15. Which is the number?

Choose one of the following answers

- ○ 3
- ○ 5
- ○ 2
- ○ 7
- ◉ No answer

Section II: Logical reasoning

- There are 20 questions in this section.
- You have **atmost 12 minutes** to answer these questions.
- **This section has a 50% chance of being chosen for payment.**
- Each question is worth 20 points. There is no penalty for not answering (No answer option). There is a penalty of -10 points per incorrect answer.

1. In the United States between 1850 and 1880, the number of farmers continued to increase, but at a rate lower than that of the general population.

Which of the following statements directly contradicts the information presented above?

Choose one of the following answers

Time remaining 00:11:39

- ○ The rate of growth of the United States labor force and the rate of growth of the general population rose simultaneously in the 30 years between 1850 and 1880.
- ○ The proportion of farmers in the United States labor force remained constant in the 30 years between 1850 and 1880.
- ○ The proportion of farmers in the United States labor force decreased from 64 percent in 1850 to 49 percent in 1880.
- ○ The proportion of farmers in the general population increased from 68 percent in 1850 to 72 percent in 1880.
- ◉ No answer

2. World Trade Organization believes in Intellectual Property and Liberalized Trade Regime. Their belief is based on the assumption that:

Choose one of the following answers

- ○ Modern economy bestows equal benefits.
- ○ Knowledge-based economy and globalization is a reality.
- ○ All countries are well suited for modern trade.
- ○ Rich countries will always help the poor countries.
- ◉ No answer

APPENDIX F: Post Task questions

Section III.Participant's characteristics

1. Full Name
2. Roll No.
3. Gender: Male/Female/Non-binary/Prefer not to say
4. Medium of Primary Schooling (i.e., Class 1 to 5): English/Other
5. Caste Category: SC/ST/OBC/General/Prefer not to say/Don't know
6. Religion: Buddhism/Christanity/Hinduism/Islam/Jainism/Sikhism/Other/Prefer not to say
7. CGPA:less than equal to 3/greater than 3 but less than equal to 4/greater than 4 but less than equal to 5/greater than 5 but less than equal to 6/greater than 6 but less than equal to 7/greater than 7 but less than equal to 8/greater than 8 but less than equal to 9/greater than 9/Prefer not to say
8. Imagine that you are playing a coin toss game. In the following table you find a list of coin tosses with different payoffs. The payoff differs in how much you could lose if the coin turns up heads. For each row you need to indicate whether you would want to toss the coin (In other words, you have to decide whether you would have played the lottery for the expected payoff or not).

	I don't want to toss the coin	I want to toss the coin
If the coin turns up heads, then you lose 50, if the coin turns up tails, you win 250	○	○
If the coin turns up heads, then you lose 100, if the coin turns up tails, you win 250	○	○
If the coin turns up heads, then you lose 150, if the coin turns up tails, you win 250	○	○
If the coin turns up heads, then you lose 200, if the coin turns up tails, you win 250	○	○
If the coin turns up heads, then you lose 250, if the coin turns up tails, you win 250	○	○
If the coin turns up heads, then you lose 300, if the coin turns up tails, you win 250	○	○

Section IV:Post Quiz Questionnaire

1. In your opinion, how many questions did you answer 'correctly' (section-wise)?

2. In your opinion, how many questions did other participants get correct on an average (section wise)?

3. Around 300 Undergraduates (UG) students from your institute are participating in this quiz. In which percentile range will you rank your performance (listed from the best performing percentile, for example, top 10% of performers will fall in 'Between 100th to 90th percentile' and lowest 10% of performers will be 'Below 10th percentile')? [i.e., performance measured by "total number of correct answers from a total of 40 questions" score distribution]

4. From approximately 300 UG students taking part in this study today, top 25% of scorers plus an equal proportion of randomly selected individuals through a lucky draw from the remaining participants will get a bonus amount of INR 100. (For example: Let's say exactly 300 UG students take this quiz, then top scoring 75 students plus another 75 students selected through a lucky draw from the remaining participants will get INR 100 as bonus)
   Suppose you find out at the time of the payment that you have INDEED received the bonus payment. This may be due to the either of the following two reasons:
   i. You scored among top 25%
   ii. You were selected through a lucky draw from the remaining participants
   To which of the above reason would you attribute your selection to? [score is in terms of "number of total correct answers out of a total of 40 questions"]

5. Do you have a preference for any particular section to be considered for the payment?

6. Please enter a valid email ID (we will contact you on this email ID for payment process)