How do beliefs and preferences over jobs affect enrollment in vocational training: Experimental evidence from India

Apurav Yash Bhatiya
Bhaskar Chakravorty
Clément Imbert
Roland Rathelot*

June 24, 2025

Abstract

We survey young job seekers in rural India to understand the determinants of enrollment in a government training program with guaranteed placement into urban jobs. Respondents are over-optimistic: they expect jobs that pay more and are closer to home than actual placement opportunities. We implement an RCT and provide them with objective information on the distribution of placement salaries or job locations. The intervention successfully corrects subjects' beliefs, which affects their decision to enroll in the program. By revealed preferences, our estimates suggest that job seekers need to be paid 50% more to work outside their home state.

JEL Codes: D83, D84, R23, O15

Keywords: urban labor markets, rural youth, vocational training, belief updating, migration costs.

*Bhatiya: University of Birmingham. Chakravorty: University of Warwick. Imbert: Sciences Po and University of Warwick. Rathelot: Institut Polytechnique de Paris (ENSAE). We thank Wiji Arulamapalam, Travis Baseler, Catia Batista, Gaurav Chiplunkar, Mattia Fracchia, Sarah Frohnweiler, Matt Lowe, David McKenzie, Kate Orkin, Nina Roussille, Ashish Shenoy, and seminar participants at Birmingham, NOVAfrica Migration and Development International Conference, DULBEA-ROA Workshop, ISI Delhi, Marseille, Newcastle, Bath, Oxford CSAE Conference 2023 and RES Annual Conference 2023 for useful comments at many stages of the projects. We also thank Kritarth Jha, Maximilian Lohnert, and Poonam Panda for their excellent research assistance. This project has been funded by EQUIP-ESRC. This project would not have been possible without the amazing collaboration of BRLPS and the Ministry of Rural Development. This paper was previously circulated with the title: "Expected Benefits and Costs of Migration for Rural Youth: Experimental Evidence from India". This trial was registered with the American Economic Association registry for randomized control trials, under the title "Mobilisation for Skill Training: Experimental Evidence from Bihar", and the trial number AEARCTR-0006003. The ethical approval was obtained from the Research Ethics Committee of the University of Warwick (HSSREC 36/17-18), and IFMR (IRB00007107; FWA00014616; IORG0005894).

1 Introduction

Large rural-urban wage gaps in developing countries suggest that rural workers lack the skills to do productive jobs in urban areas, and/or are held back by job search and migration frictions, with large negative effects on aggregate productivity (Young, 2013; Gollin et al., 2014; Bryan and Morten, 2019; Tombe and Zhu, 2019). Vocational training holds the promise of equipping rural workers with the required skills and facilitating their placement into higher-paid jobs located in areas where these jobs are more abundant. However, the take-up of training schemes remains low and drop-out rates high, which limits the effectiveness of these policies (McKenzie, 2017). One possible reason for this low take-up and high drop-out could be that young job seekers dislike the high-paying jobs offered to them because of non-pecuniary amenities (Blattman and Dercon, 2018; Imbert and Papp, 2020).

In this paper, we take advantage of a vocational training program, India's DDU-GKY (Deen Dayal Upadhyaya Grameen Kaushalya Yojana), which trains young rural workers and guarantees them placement in a formal urban job for free. In addition to providing new skills, the program alleviates most barriers to accessing better jobs. However, the program suffers from high drop-out: only two-thirds of its 1.7 million trainees so far have taken the job offered to them, and 54% stayed in it for three months.[1] This suggests that many trainees do not know what the placement jobs really are, or do not want them once offered. To investigate the role of beliefs and preferences, we carried out a survey and a field experiment with prospective DDU-GKY trainees. The survey reveals that candidates think placement jobs pay more and are closer to home than in reality. We experimentally correct these beliefs by providing factual information on placement jobs' location, salary, or both. The relative reductions in enrollment due to changes in beliefs on location and salary suggest that strong home location preferences are a major barrier to job take-up.

Specifically, we surveyed 876 rural youth from Bihar (India) who attended 63 "mobilization" camps where prospective trainees learned about DDU-GKY from training provider and government representatives. The survey suggests that the average candidate held over-optimistic expectations about placement opportunities: they expected 55% of jobs to be in their home state (the signal is 20%) and the average wage to be Rs. 9,800 (the signal is about Rs. 8,300). This may be due to self-selection of over-optimistic candidates into the camps, but "mobilizers" also had incentives to encourage over-optimistic beliefs in order to enroll more trainees. In any case, in our context, information frictions made rural young workers more willing to enroll in the program, but also more likely to drop out once they learned about actual working conditions in the future.

[1] Official statistics from http://DDU-GKY.gov.in/ accessed on 29th July 2024

We then provided information on the distribution of jobs provided by the program in the last year in terms of location (in/out of state) and salary (in 5 bins). Our intervention was successful in reducing the gap between beliefs and reality: posterior beliefs in the treated group were closer to the signal and significantly different from their own priors and from posteriors in the control group. Our intervention corrected 62% of the initial bias on the job location and 90% of the bias on average salary. Belief updating was persistent: treated individuals held on to the updated beliefs up to four weeks after the intervention. We check that respondents adjusted their expectations about their own career if they completed the training, and not their expectations about outside options if they did not enroll.

Finally, we match the survey sample with administrative data on training enrollment and estimate the effect of salary and location expectations, instrumented by treatment assignment, on the decision to enroll. We find that the decrease in salary expectations and in the perceived likelihood of finding a job in their home state made the average treated candidate less likely to take part in the training program overall. The relative effect of location and wage expectations on the decision to enroll also provides revealed preference estimates of the perceived cost to move out of state. We estimate that rural job seekers require a salary that is 50% higher to take up a job out of their home state, which is much higher than the premium actually offered in the placement jobs (only 3%). Comparing our estimates to those in the literature, we find that they are lower than Tombe and Zhu (2019)'s structurally estimated migration costs in China, which are twice as large across provinces as within ($0.97 vs. $0.45). Jobs out of state are located on average 10 times further away, which implies an elasticity of migration costs to distance of 5%, and places our estimates between Bryan and Morten (2019)'s for Indonesia (15%) and the US (2%). It is not surprising that our estimates of migration costs, based on a subpopulation of young rural job seekers who expressed an interest in skilled jobs, are lower than the population-wide estimates in the literature. But even in that subpopulation, migration costs are substantial and large enough that they would give up valuable placement opportunities.

Our paper relates to four strands of the literature. First, our paper adds to the literature that studies job search frictions and barriers to youth unemployment in developing countries (see, McKenzie (2017) for a review). Existing research highlights the importance of training (Alfonsi et al., 2020; Adhvaryu et al., 2023), of signaling one's skills (Carranza et al., 2020; Bassi and Nansamba, 2022), of search costs (Franklin, 2018; Abebe et al., 2021a,b), and information frictions (Hicks et al., 2011; Jensen, 2012). Recent contributions highlight the role of job seekers' often misplaced expectations about their labor market prospects to interpret the effect of experimental interventions aimed at improving their employment outcomes (Abebe et al., 2017; Banerjee and Sequeira, 2020; Alfonsi et al., 2022; Bandiera et al., 2023). We study a context in which most job search frictions are alleviated by the offer of a free training and placement program (DDU-GKY), which allows us to focus on the role of job seekers' beliefs and preferences.

Two other papers study the role of DDU-GKY trainees' preferences on their labor market outcomes. Banerjee and Chiplunkar (2022) inform placement officers about trainees' preferences regarding their placement jobs and find that it leads to a better match and retention in the program. In Chakravorty et al. (2024), we show that informing trainees about placement jobs improves retention, presumably by inducing self-selection of trainees who are a better fit for the available jobs. Our contribution in this paper is to precisely measure and experimentally manipulate the beliefs of prospective trainees through an information treatment and to estimate the causal effect of beliefs on labor market decisions. We show that prospective candidates hold over-optimistic beliefs about the location and the pay of placement jobs and that correcting these beliefs reduces enrollment in the program. Our results suggest that location preferences are important barriers for rural job seekers to access formal (urban) jobs.

Second, there is a related and abundant literature on job search frictions in developed countries (Altmann et al., 2018; Belot et al., 2019, 2021; Kircher, 2022). Like the literature in developing countries, it includes structural work highlighting the importance of spatial frictions in job search (Van Ommeren and Fosgerau, 2009; Manning and Petrongolo, 2017; Marinescu and Rathelot, 2018; Schmutz and Sidibé, 2019). This strand of literature also includes lab-in-the-field experiments that estimate the value of non-monetary job amenities, such as commuting time (Mas and Pallais, 2017). Using an experimental design close to ours, Cullen and Perez-Truglia (2022) test the effect of information about pay inequality on employee motivation, and Jäger et al. (2022) the effect of information about outside options on job search intentions. To our knowledge, we are the first to use an experimental information treatment to quantify the value of job location in a real-world context.

Third, we contribute to the literature that aims to understand the sources of rural-urban wage gaps in developing countries (Gollin et al., 2014; Bryan and Morten, 2019; Tombe and Zhu, 2019). The literature emphasizes the lack of skills among rural workers (Young, 2013), financial constraints, and uninsured risk (Bryan et al., 2014; Munshi and Rosenzweig, 2016; Meghir et al., 2022). By contrast, we work in a context where skill mismatch, financial constraints, and risk are minimized by the offer of a free vocational training program with guaranteed placement. Instead, we focus on the role of rural job seekers' beliefs about urban jobs and their preferences about salary and location. A paper close to ours is Baseler (2022), who shows that rural workers in Kenya underestimate urban wages and that experimentally providing accurate information increases migration to the capital city. In a similar vein, Frohnweiler et al. (2022) experimentally provide information on regional income differentials in Ghana and Uganda and find that it affects the destination choices but not the intention to migrate. In our setting, prospective candidates are on average over-optimistic about the urban placement jobs, so that accurate information reduces their willingness to join the program. We are the first to use experimental variation in beliefs about salary and location to provide revealed preference estimates of migration costs. We find that rural young job seekers require 50% higher salary for a job located outside of their home state. Qualitatively, our findings resonate with Kone et al. (2018)'s, who document substantial inter-state migration barriers in India, and with Imbert and Papp (2020) and Lagakos et al. (2023)'s findings of high non-monetary costs of seasonal migration in India and Bangladesh respectively. Quantitatively, our estimates of migration costs are lower than those from Bryan and Morten (2019) for Indonesia and Tombe and Zhu (2019) for China, which is likely due to the fact that we focus on young job seekers. Even then, migration costs are high enough to prevent them from taking up formal urban jobs.

Finally, the literature which is most closely related to ours in terms of design are recent lab-in-the-field and field experiments that study the determinants of international migration decisions (Shrestha, 2020; Bah and Batista, 2018; Batista and McKenzie, 2021; Bazzi et al., 2021; Bah et al., 2022).[2] On the one hand, Bah and Batista (2018) and Batista and McKenzie (2021) study the determinants of international migration intentions in a lab-in-the-field setting, with only reported migration intentions as an outcome. On the other, recent field experiments provide information on different aspects of the migration experience (e.g. intermediaries, mortality risk) and assess their effect on migration decisions without precisely identifying beliefs or preferences (Bazzi et al., 2021; Bah et al., 2022). One exception is Shrestha (2020)'s, who experimentally provides information on earnings abroad and on the probability of dying to potential international migrants in Nepal. He estimates the effect of beliefs about earnings and mortality risk on international migration decisions to compute the value of a statistical life. Like his, our design combines the advantage of a lab-in-the-field setting, by precise measurement of belief updating, with the advantage of a field experiment, enabling us to look at the real-world decision to be trained and placed in urban areas. We are the first to use this design to estimate location preferences among potential rural-urban migrants.

[2]The importance of beliefs in international migration has been emphasized since at least McKenzie et al. (2013), and recently by McKenzie and Yang (2022) in their recent literature survey.

2 Context and Experimental design

2.1 Context

India, like other developing countries, has large spatial differences in rural-urban wages. A cross-national comparison of internal migration by Bell et al. (2015) shows that India has among the lowest internal migration rate. In 2014, the Ministry of Rural Development (MoRD) launched the "Deen Dayan Upadhyaya Grameen Kaushal Yojana" (DDU-GKY) to tackle this challenge. DDU-GKY program is a residential training and placement program that targets unemployed rural youth aged 15-35 years from poor families and places them in jobs outside the home state (often in Delhi, Tamil Nadu and Kerala).[3] The program focuses on rural youth, with mandatory coverage of females and socially disadvantaged groups. As compared to training-only programs, DDU-GKY shifts the emphasis to training and placement, with a mandatory placement of at least 70% candidates.

MoRD provides overall monitoring, policy formulation and funding to DDU-GKY is in collaboration with the states (75% central funding and 25% state funding). The scheme is implemented in the public-private partnership (PPP) model, where registered private training partners (PIAs) are responsible for mobilizing the candidates, provide training and placement in salaried jobs, targeting rural youths from poor families. PIAs are also responsible for providing post-placement support to the trainees and ensure retention in the placed jobs. The total investment includes: (1) training costs: covering classroom instruction, materials, sector-specific training, and soft skills development; (2) residential facilities: providing food and accommodation for trainees for the training duration; (3) post-placement support: financial assistance of Rs 1000 per month for 2-6 months, depending on placement location. The funds are disbursed in phases based on enrollment, completion, and placement milestones.

The pool of participants attending mobilization camps typically consists of rural youth from lower-income families, many of whom have limited employment options and little exposure to formal training programs. The opportunity cost of training includes foregone earnings, as many rural youth must give up informal or agricultural work during the 3–12 month program period. Additionally, uncertainty about job placement discourages participation, as prior research (Chakravorty et al., 2024) shows high dropout rates even after placement, suggesting that many trainees ultimately find urban jobs unsuitable. Even for those who do not take up a job, the program offers benefits such as skill development,

[3]The training is a combination of classroom and on-the-job learning. It has two components: soft skills, English, and IT training, followed by sector-specific instruction. Courses vary in duration from 3 to 12 months, with on-the-job training lasting between 30 to 120 days, depending on course length.

which can improve future job prospects, and certification, which may be valuable in local labor markets or for self-employment. Our earlier work (Chakravorty et al., 2023) found that DDU-GKY graduates from Bihar and Jharkhand had better employment outcomes than dropouts, highlighting the program's role in fostering employment resilience during economic downturns.

For the purpose of this study, we collaborated with the Bihar Rural Livelihood Promotion Society (BRLPS), who is in charge of DDU-GKY in the state of Bihar. We worked in "mobilization camps" organized by BRLPS in collaboration with the private partners in charge of training and placement (called Project Implementing Agencies- PIAs). Mobilization camps serve as the primary mechanism for recruiting candidates into the DDU-GKY training program. Prospective candidates learn about mobilization camps through various outreach channels, including village networks led by Job Resource Persons (JRPs), printed materials in public spaces, door-to-door visits by PIA mobilizers, word of mouth from past trainees and community leaders, and local events or job fairs showcasing training opportunities.

Mobilization camps are typically held at accessible community locations such as schools, village meeting halls, private community centers, Gram Panchayat offices, or block headquarters depending on local availability to ensure ease of participation for rural youth and to maximize attendance while keeping travel costs minimal. The time prospective candidates spend at these camps varies, but in a typical session, candidates spend around 1 hour engaging with program representatives.

The structure of a mobilization camp generally includes an introduction to the program by the JRP who provides overview of the scheme, its benefits, and eligibility criteria. Next, the PIA mobilizers from different training centers present details about their courses, facilities, job prospects, and placement locations. Candidates interested in specific training programs can ask questions, clarify doubts, and receive guidance on available opportunities. Those who express interest in enrolling are guided to visit the training centers with their parents for the next steps of the registration, including eligibility verification and document submission.

From qualitative interviews, we learned that potential trainees were misinformed about DDU-GKY placement opportunities, i.e. that they overestimated the wages offered and underestimated how far the jobs were. We suspected that this misinformation could stem from mobilizers and JRP themselves, who have professional incentives to enroll the maximum number of candidates. The mismatch between trainees' expectations and placement opportunities contributes to high drop-out rates, a major concern for BRLPS.[4]

[4] In Chakravorty et al. (2024), we document that 88% of enrolled candidates complete training, but only

2.2 Intervention

We designed an information intervention to correct the labor market expectations of the potential trainees. At the end of the mobilization camp, we invited candidates to answer a few questions from the survey team. In an ideal setting, we would collect information on expectations and beliefs prior to recruitment or announcement of the mobilization camp by the JRP. In practice, however, identifying and surveying individuals who are eligible for the program and are interested in attending a mobilization camp poses significant logistical and financial challenges. Moreover, given that the mobilization process is itself designed to encourage participation, there is likely to be strong self-selection among those who choose to attend, making it difficult to disentangle whether the source of misperception is external misinformation or internal biases.

Our survey measured candidates' priors on DDU-GKY jobs' location and salaries (see below). After these questions, respondents were randomly assigned to one of the four intervention arms (individual-level randomization). In the control group, the candidates watched a basic informational video about the DDU-GKY program, the training center, accommodation and food facilities, and classrooms. In the video, two past beneficiaries described their (positive) experience with DDU-GKY. The control video did not provide any information on job location or wages offered. In the location treatment group, candidates watched the basic information video and one additional video that provided information on the distribution of placement job location for past DDU-GKY candidates. Similarly, in the salary treatment group, the second video showed the distribution of salaries of past placement jobs. In the salary treatment $\times$ location treatment group, the candidates watched all three videos.

Specifically, the intervention videos displayed 10 candidates who were allocated into two bins for the location treatment (inside state and outside state) and five bins for the salary treatment (less than Rs 6000 per month, Rs 6000 - 8000 per month, Rs 8000 - 10000 per month, Rs 10000 - Rs 12000 per month and more than Rs 12000 per month). Since the wages and job offers differ across male and female candidates (primarily due to different training sectors), the distributions were tailored to the gender of the candidate.[5] The distribution of wages and location for the placement job was obtained from a parallel project carried out the same year and in the same state (Chakravorty et al., 2024). Administrative data have incomplete information on the placement jobs of the candidates as PIAs don't focus

45% join their placement job, and 33% are in their placement job after five months. We show that providing information about placement jobs to trainees has no effect on training completion or placement, but improves retention conditional on placement, which we interpret as evidence of improved self-selection into placement.

[5] Appendix Figure A1 - A4 show snippets of the location and salary intervention videos for females and males, respectively. Appendix Section B provides a detailed transcript of each video.

on tracking candidates once they have left the training center. By contrast, surveys from Chakravorty et al. (2024) followed a sample of 2,488 DDU-GKY trainees from enrollment to five months after training completion with an attrition below 5%.

At the end of the mobilization activity and the surveys, candidates immediately left the camp to return home, limiting the opportunity for interactions within the camp setting itself. However, individuals from the same panchayat who attended the same camp may have commuted together, interacted, and shared information afterward. Hence, we define spillover exposure at the mobilization camp $\times$ panchayat level.[6] Our hypothesis is that that if spillovers occur, respondents might develop distrust toward any information provided by 'outsiders' (the research team) and instead place greater trust in the information from the job resource person from their local community.

2.3 Data

Our research relies on primary data collected from three rounds of surveys.[7] In addition, we used administrative data, which we matched with the survey data.

The baseline survey was administered to all participants in the mobilization camps after the trainees had received information from the JRP and/or the PIA mobilizer. It was a face-to-face interview with individual trainees between mid-December 2019 and mid-February 2020. The baseline questionnaire first collected information about the probability of enrolling in the training and about their priors on the distribution of wages and location of DDU-GKY jobs. Specifically, the survey asked 'After the training, if 10 people like you get a job. How many will get a job inside of Bihar, and how many will get a job outside of Bihar?" and 'After the training, if 10 people like you get a job. How many will get a job with a monthly salary of less than Rs 6000 / Rs 6000 - 8000 / Rs 8000 - 10000 / Rs 10000 - Rs 12000 / more than Rs 12000 per month'. To make it easier for respondents, we followed best practices from Delavande et al. (2011), and gave them ten marbles which we asked them to distribute into cups (one for each option). Then the survey provided information on the location and earnings distribution of DDU-GKY jobs following the randomized treatment assignment and customized depending on the gender of the candidate. Finally, the survey measured posterior beliefs about wages and job location, following the same methodology as for the priors. In addition, it asked about the posterior probability to enroll in the training, expected earnings in a year if they completed the training, counterfactual earnings if

[6] This was pre-registered in the pre-analysis plan.

[7] Appendix Section C provides the consent forms associated with the baseline and follow-up surveys.

they did not, and socio-economic characteristics.

The two follow-up surveys were conducted on the phone with the trainees one week and four weeks after the baseline survey for all respondents. Qualitative interviews with JRPs and PIAs informed us that most candidates who want to enroll on the training program enroll within a week or 10 days of the mobilization camp. The objective of these surveys was to collect information about the posterior beliefs on wages and job location, expected and counterfactual earnings at the time when the candidates were making a decision to enroll in the program. The surveys also asked whether the candidate had visited the training center or enrolled in the program. We could not follow the respondents' journey through the training and beyond as the training centers were shut down due to the COVID-19 pandemic towards the end of March 2020, however, respondents from the last round of baseline surveys had enough time to enroll.

The administrative data comes from the management information system (MIS) of BRLPS and was compiled from the PIAs report to the state administration. This dataset was obtained in July 2020 and includes official information on candidate enrollment for the last 2 years. We matched it to the survey dataset by mobile number, name and district of the candidate.

2.4 Summary statistics and balance tests

Our sample includes 876 candidates from 63 mobilization camps organized in Bihar.[8] The surveys were conducted between December 2019 and February 2020.[9] Information from the camp activity survey suggests that 74% of the camps were attended by the PIA mobilizer. All camps had the presence of a JRP. In 9.5% of the camps (6 out of 63), neither the JRP nor the mobilizer provided an introduction to the program. In 30% of the camps (19 out of 63), both the JRP and mobilizer spoke about the DDU-GKY program.

The summary statistics of our baseline variables are provided in Appendix Table A1. The average age of candidates in our sample was 20, and almost 58% were females. In terms of social category, 30% of the candidates came from the Scheduled Castes and Scheduled Tribes, and 55% were OBCs, which shows the pro-poor targeting of the DDU-GKY program.

[8]Our total survey sample was 880. However, in 4 camps there was only 1 candidate each. We exclude these camps from our analysis. Correia (2015) suggests that singleton observations together with mobilization camp fixed effects can overstate the statistical significance and lead to incorrect inference.

[9]The COVID-19 lockdowns were introduced in India towards the end of March 2020 and are unlikely to have affected the mobilization camps and the candidate's decision to enroll on the program.

Both females and males say it would not be difficult for their family if they enroll in the training program and that there is almost 80% probability of enrolling in the program. This suggests JRPs target the candidates well: candidates who fulfill the program targeting and those who are eager to take part in the training were present in the mobilization camps. Balancing tests suggest that there were no issues with the randomization (Appendix Table A2). The attrition rate in both follow-up rounds is low (almost 6%) and similar across all treatment and control groups (Appendix Table A5).

Figure A5 shows the misperceptions in labor market beliefs. We measure misperceptions by comparing the prior beliefs with the signal. Less than 5% of the respondents' prior beliefs for the location fell within $\pm$ 5% of the signal. The majority of the respondents underestimated the number of candidates outside state, often by a large margin: the mean absolute error was 50%. On the average salary, the mean absolute error was 25%, only 12% of candidates' prior beliefs were within 5% of the signal, and a majority of the candidates overestimated the average salary.

These misperceptions are shaped by multiple factors. Since the baseline survey was conducted after the mobilization camps took place, beliefs reflect a combination of prior misconceptions (unrelated to the camp), information provided by job-resource persons (JRPs) who recruited individuals, and reinforcement by mobilizers during the camps. This raises an important distinction between correcting misinformation—where inaccurate information is deliberately provided—and correcting inaccurate beliefs, which arise from prior misconceptions, selective exposure to information, and cognitive biases. To better understand the sources of these misperceptions, we conducted additional analyses.

First, as shown in Table A3, we find no significant differences in misperceptions across districts where the camps were organized or by the presence of a mobilizer in the camp (notably, 26% of the camps did not have a mobilizer present). If mobilizers were systematically inflating expectations during the camp, we would expect to see greater misperceptions in locations where mobilizers were actively engaged. The absence of such differences suggests that belief inaccuracies are not primarily driven by mobilizers within the camps. However, it remains possible that JRPs (who operate across districts) played a role in shaping expectations before the camps, meaning that the mobilization process itself may not be the primary driver of belief inaccuracies—rather, the recruitment phase may have already set a common level of misperceptions.

Second, Table A4 highlights that misperceptions in prior beliefs are systematically related to socio-demographic characteristics. Notably, female participants exhibit significantly lower misperceptions about job location but significantly higher misperceptions about salary, and these effects are both large and statistically significant. Appendix Figure A6 further illustrates gender differences in salary misperceptions through a density plot of prior salary beliefs. While male and female respondents had similar salary expectations (Rs 9,800 on average), actual DDU-GKY placement salaries were lower for women (Rs 7,600) than for men (Rs 9,000). One potential explanation is that women base their expectations on the experiences of male migrants, whereas female-dominated sectors, such as garment factories, tend to offer lower wages. Higher education and belonging to the OBC social category are also associated with greater misperceptions about job location, but these characteristics do not significantly affect salary misperceptions. These results change marginally as we exclude or include camp fixed effects, suggesting that belief formation is not entirely driven by external information but also by internal cognitive biases, preferences, and differential access to informal networks that may influence expectations about urban employment opportunities. This motivated our decision to tailor information interventions based on gender.

3 Empirical Framework

3.1 Beliefs

Our empirical analysis largely follows our pre-registered pre-analysis plan available online[10] and in the Appendix Section D. We explain the deviations from the pre-analysis plan in Appendix Section E and highlight them here. Our empirical model follows the standard belief-updating framework in the literature (Fuster and Zafar, 2023; Jones and Santos, 2022), where posterior beliefs are modeled as a function of prior beliefs and new information. We estimate the effect of our intervention on labor market beliefs regarding the location $j = l$ or the salary $j = s$ of DDU-GKY placement jobs for individual i present in mobilization camp c using the following specification:

$$Posterior_{ic}^{j} - Prior_{ic}^{j} = \gamma_j T_{ic}^{j} + X_{ic}'\alpha + \delta_c + \varepsilon_{ic} \tag{1}$$

$Prior_{ic}^{j}$ and $Posterior_{ic}^{j}$ denote the respondent i's prior and posterior distributions for DDU-GKY placement jobs' salary and location. Prior distributions are measured by the baseline survey before the intervention. Posterior distributions are measured either by the baseline survey after the intervention or by the two follow-up surveys. Location beliefs are measured as the number of trainees (out of 10) who get a job outside of Bihar. Salary

[10] American Economic Association registry for randomized control trials, under the title "Mobilisation for Skill Training: Experimental Evidence from Bihar", and the trial number AEARCTR-000600.

beliefs are measured as the average expected salary, computed as the sum of the mean salary in each bin times the share of candidates (out of 10) assigned to each bin.[11] T_{ic}^{j} is an indicator variable equal to one if the candidate i received information about salary ($j = s$) or location ($j = l$). The coefficient of interest γ_j is the estimate of how treated individuals update their labor market beliefs on average as compared to those who did not receive the treatment j. δ_c are mobilization camps fixed effects, and X_i denotes a vector of individual characteristics selected using a post-double-selection lasso (Belloni et al., 2014). Standard errors are clustered at the mobilization camp level.

The sufficiency assumption underlying this approach is that individuals' prior beliefs fully encapsulate all information available before the intervention, and any observed changes in beliefs are directly attributable to the treatment. Moreover, we assess the invariance assumption by testing whether receiving information about job location influences beliefs about salary and vice versa. If invariance holds, individuals should update only in response to relevant information. However, if belief updating spills over across dimensions, this would indicate a violation of invariance, suggesting that individuals integrate unrelated information into their belief revision. To test this, we regress changes in beliefs on both treatment dummies and their interaction term, which allows us to detect whether cross-treatment effects exist.

Our model implicitly accounts for stability in belief updating over time. Specifically, if individuals retain the new information provided during the mobilization camp, their updated beliefs should persist across survey rounds. Alternatively, if they partially revert towards their priors, this would suggest an adjustment process rather than a one-time shift. Our follow-up surveys allow us to examine whether individuals' revised beliefs remain stable or exhibit further revision, providing additional insights into the dynamics of learning and information retention in this context.

To gain deeper insights into belief changes, we examine whether the information intervention altered respondents' perceptions of labor market conditions generally or specifically influenced their expectations about their own career trajectories. In particular, while the intervention provided new information about the distribution of salaries and job locations in DDU-GKY placements, it is unclear whether respondents updated their expectations about the labor market as a whole or only about their own future if they enrolled. To test this, we use survey questions that capture respondents' expected earnings and job location in a year under two scenarios: if they enrolled in DDU-GKY and if they did not. We use specification 1 to verify that the intervention altered respondents' expectations regard-

[11] We use Rs 5,000 as the mean salary in the "less than Rs 6,000" bin and Rs 13,000 as the mean salary in the "more than Rs 12000" bin. The appendix presents results using the median salary as a robustness check.

ing their own outcomes if they joined the program, without significantly changing their broader labor market expectations outside of it.

3.2 Enrollment

Our goal is to understand how labor market expectations affect individuals' decisions to enroll in the training program. For this, we use a 2SLS estimation procedure similar to the one used by Cullen and Perez-Truglia (2022) and Jäger et al. (2022) in other contexts. Specifically, we instrument changes in beliefs in each dimension $j \in \{l, s\}$ $Posterior_{ic}^{j} - Prior_{ic}^{j}$ with the treatment indicators T_i^j, $Signal^j - Prior_{ic}^j$, which measures how far their priors were from the signal, and the interaction between them.[12] The first stages write:

$$\begin{aligned} Posterior_{ic}^{j} - Prior_{ic}^{j} = {} & \beta_1^l T_{ic}^l + \beta_2^l (Signal^l - Prior_{ic}^l) + \beta_3^l (Signal^l - Prior_{ic}^l) \times T_{ic}^l \\ & + \beta_1^s T_{ic}^s + \beta_2^s (Signal^s - Prior_{ic}^s) + \beta_3^s (Signal^s - Prior_{ic}^s) \times T_{ic}^s \\ & + X_{ic}'\alpha + \delta_c + \varepsilon_{ic} \quad j \in \{l, s\} \end{aligned} \tag{2}$$

The coefficient β_1^j captures any level shift in beliefs due to each treatment and β_3^j captures any differential updating by individuals whose beliefs were further away from the signal. This specification follows theoretical models of Bayesian updating (Fuster and Zafar, 2023), where individuals incorporate new information in proportion to its credibility and their initial distance from the signal. The β_3^l and β_3^s coefficients correspond to these weights in the updating process.

The signal was derived from a parallel project conducted in the same year and state (Chakravorty et al., 2024). Specifically, the location signal represents the share of trainees who secured jobs outside Bihar, rounded and rescaled to a 1-10 scale. The salary signal reflects the average monthly earnings of respondents in salaried employment. Given that wages and job offers vary significantly by gender due to different training sectors, we tailored the signals accordingly for male and female candidates. The same signal is applied to individuals regardless of their treatment status.

The outcome is the difference between $I(Enrollment)_{ic}^{Posterior}$, a dummy variable for program enrollment, and $P(Enrollment)_{ic}^{Prior}$, the expected probability to enroll in the baseline

[12] We deviated from the pre-analysis plan here by estimating a single first-stage equation that jointly includes both treatment arms rather than a separate first-stage equations for each treatment dimension. This allows us to account for interactions between treatment dimensions and improves estimation efficiency.

survey prior to the intervention.[13] The second stage of the estimation is:

$$I(Enrollment)_{ic}^{Posterior} - P(Enrollment)_{ic}^{Prior} = \beta_l(Posterior_{ic}^{l} - Prior_{ic}^{l}) + \beta_s(Posterior_{ic}^{s} - Prior_{ic}^{s}) + X_{ic}'\alpha + \delta_c + \varepsilon_{ic} \quad (3)$$

Both stages include mobilization camps fixed effects (δ_c), and a vector of individual characteristics (X_i) selected using a post-double-selection lasso (Belloni et al., 2014).

4 Results

4.1 Labor Market Beliefs

Figure 1a and 1b display graphically how the treatments changed labor market beliefs. Prior to the intervention, there is no difference in beliefs between the treatment and the control group. Respondents are over-optimistic: regarding location, they believe that a majority (55%) of placement jobs are in the state, when the signal is less than 20%, and regarding the salary, they believe that half of the jobs pay more than 10k, when the signal is less than 10%. After receiving the signal, the treatment group revise their expectations downward: they now attribute a 28% probability of having a job in the state, and a 30% probability of earning more than 10k. Beliefs do not change in the control.

Table 1 presents the effects of the information intervention on changes in beliefs (posterior − prior) about job location (Panel A) and average salary (Panel B), estimated using specification 1. To test whether the change in beliefs were persistent, we consider posterior beliefs at three different times: in the baseline survey just after the intervention (Column 1), one week after the intervention (Columns 2: Followup 1w), and four weeks after the intervention (Column 3: Followup 4w). The number of observations changes slightly across columns due to attrition. As Panel A in Table 1 shows, the control group who believed that 42% of placement jobs were outside the state at baseline barely updated their belief during the baseline survey (+5pp.) and in the following four weeks (+8pp.). By contrast, as Figure 1a showed, the treatment group updated their belief strongly upward (+25pp.) during the survey. One week after the survey, only half of this update remained (13%), but it had not decayed further four weeks later (12%). Table 1 Panel B turns to average salary expectations (in Rs 1,000). The control group at baseline believed that the placement jobs on average paid Rs. 9,873 and did not update their prior at all in the course of the following

[13] As compared to our pre-analysis plan, we refined the second-stage outcome by modeling changes in enrollment behavior relative to prior expectations. This approach provides a clearer measure of how belief updating influences decision-making, rather than focusing solely on absolute enrollment rates.

weeks. By contrast, the treatment group revised downwards their salary expectations by (Rs -1,463) during the survey, and again about half of that change was present a week later (Rs -655), with almost no decay four weeks later (Rs -633). These results take the average salary as an outcome, but the information provided was about the whole salary distribution: in Appendix Table A6, we check that the treatment also shifted the median closer to the signal and reduced the variance of salary expectations.

The results so far focus on respondents' beliefs about the distribution of placement jobs, but it could be that the intervention did not change their expectations about what would happen to them personally if they enrolled in the program. We check this by using as outcome respondents' expectations about where they would be and how much they would earn if they completed the training program. As Table 2 Panel A and C show, respondents in the control group are even more optimistic about their own prospect a year after training than the average placement job: only 34% believe they will be out of state, and on average they expect to earn over Rs 13,000. Reassuringly, respondents in the treatment group become less optimistic, with an increase by 7pp. in the probability to be out of state, and a Rs 1,700 reduction in their expected wage. Over the course of the following four weeks, the effects strengthened for location (+10pp.) and weakened for salary (Rs -1,100).

Another important question is whether the information treatment changed their overall labor market outlook, including the jobs they could get outside of the training program. We investigate this using as outcome respondents' expectations about where they would be and how much they would earn if the did not complete the program As (Table 2 Panel B and D). Interestingly, respondents in the control group do not generally expect to migrate out of state (between 8 and 11% depending on the survey), and their salary expectations are low (between Rs 6,358 and 7,470 depending on the survey). These low salary expectations may in fact be accurate: in a companion project in the same context, we find that the salary of respondents who enroll but drop out of training is Rs 7,600 (Chakravorty et al., 2024). Reassuringly, the information treatment has no effect on respondents' expectations about their location or earnings if they did not complete the training.

Following our pre-analysis plan, we extend our analysis of belief updating in three more ways. First, we test whether information about salary changed beliefs about location and vice-versa: Appendix Table A7 suggests that there is no evidence of cross-treatment or interaction effects. Second, we consider belief updating separately by gender, caste and education level: as Appendix Figure A7 shows, all groups updated their beliefs on location, but the update on salary was stronger for female and less educated respondents. Third, we test whether the treatment group may have affected beliefs of the control group by telling them about the information they received. To identify spillovers, we measure the fraction of peers (candidates from the same panchayat who attended the same mobilization camp) who were treated. Column 2 in Appendix Table A8 presents some evidence that in the first follow-up spillovers made control individuals more optimistic as compared to their baseline posterior (i.e. made them deviate more from the truth) about location (Panel A), but not about salary (Panel B). In contrast treated individuals' beliefs were not affected by their peers' treatment. There is also no evidence of spillovers in the four weeks follow up.[14] On the whole, we find little evidence that control individuals learn from their treated peers.

4.2 Enrollment and Preferences

We now examine whether the change in respondents' beliefs about the placement jobs and their own labor market prospects if they join the program actually influenced their decision to enroll. Table 3 presents the estimates of the first stage. The estimate of β_3^l in Columns 1 and 2 suggest that respondents who underestimated the probability of being placed out of state more also updated their beliefs more. Specifically, respondents whose prior were 10pp. below the signal (the average respondent is 35pp. below the signal) updated by 2.2pp. after one week and 2.4pp. after four weeks (both highly significant). Similarly, the estimates of β_3^s in Columns 3 and 4 suggest that the salary treatment had a stronger effect on the beliefs of respondents whose priors were further away from the signal: respondents who were Rs 1,000 below the signal (the average respondent was about Rs 2,000 below) updated their beliefs by Rs 137 (Rs 182 after four weeks).

In addition, the significant coefficients on β_2^j suggest that individuals—regardless of treatment status—substantially adjust their beliefs in response to the gap between their prior beliefs and the signal. This indicates that belief updating is not solely driven by direct treatment but also occurs more broadly due to multiple sources of information. One factor contributing to this finding is that participants likely receive labor market information after the mobilization camp from various channels beyond the intervention itself. Individuals may discuss job prospects, salaries, and locations with family members, particularly parents and siblings, when deciding whether to enroll in the training program. These discussions could reinforce or modify their beliefs, leading to adjustments even among those not directly treated with the salary or location intervention, consistent with models of social learning.

Additionally, all participants—including those in the control group—were shown a basic informational video about the DDU-GKY program, covering details about the training

[14] In Appendix Table A12 we test separately the effect of spillovers on priors, baseline and follow-up posteriors (this analysis was not pre-specified). Reassuringly we find no spillovers during the baseline survey, i.e. an individual's prior or posterior beliefs did not depend on whether their peers were treated.

center, accommodation, food, and classroom experience. While this video did not provide explicit information about job salaries or locations, it featured two past beneficiaries sharing positive experiences. Even without direct salary or location details, this information video may have subtly shaped participants' perceptions of labor market prospects, particularly for those whose priors were far from the actual signal. Individuals with more extreme priors may have been especially prone to updating, as they were the ones likely holding the most inaccurate or uncertain beliefs before receiving any program-related information.

Table 4 presents the estimates for the second stage, i.e. the effect of beliefs about placement jobs on the decision to enroll in the program.[15] We present results using as outcome either the difference between self-declared enrollment in the two follow-up surveys and intentions to enroll at baseline (Columns 1 and 2) or the difference between enrollment in the administrative data and intentions to enroll at baseline (Column 3). In the control group, intentions to enroll at baseline were high (79%), much higher than the self-reported measures of enrollment (19% after a week and 24% after four weeks), which were themselves higher than the enrollment rate confirmed in the administrative data (10%). These discrepancies were likely due to a combination of actual barriers to enrollment (e.g. availability of training, eligibility criteria, parental opposition, etc.) and experimenter demand effect: respondents likely anticipated that researchers expected them to enroll in the program.

Reassuringly, the estimated effects of beliefs on enrollment decisions are very similar across data sources, with more statistical precision for the more reliable administrative measure. The estimates in Table 4 Column 3 suggest that a 10pp decrease in the probability of being placed out of state would increase enrollment by 1.2pp (about 12%) and that a Rs 1,000 higher salary would increase enrollment by 2.4pp (about 24%).[16] Since the expected probability of migrating out of state increased by 12pp and the expected salary decreased by Rs 633, these estimates suggest that the location and the salary treatment reduced enrollment by about 1.4pp each.[17]

To understand how survey respondents trade off salary and location, we interpret the ratio of the two coefficients (salary/location). We perform 200 replicates of the mobilization camps with replacement and report the mean of this ratio and the 95% confidence interval at the bottom of the table. We find that the ratio is 2 and is close to the bootstrapped ratio

[15] The Kleibergen-Paap F-stat for the joint significance of the two instruments and the Sanderson-Windmeijer partial F-stat for the instruments' joint significance in the two separate first-stage regressions all suggest that the instruments are strong.

[16] Following the pre-analysis plan, we estimate the effect of labor market beliefs on intentions to enroll in the program. Intentions to enroll declined over time, but at four weeks they were still much higher than even self-reported enrollment (60% as compared to 24%). We find no effect of beliefs on intentions (Table A9).

[17] We deviate from the pre-analysis plan and estimate the effect of receiving any information treatment on enrollment: the estimate is large (-3.9pp or -40%) but insignificant (Table A10).

(2.14 in Column 3). The location beliefs are expressed in 10% probability of being outside the state and the salary beliefs expressed in Rs 1,000. Hence the 2 ratio implies that in order to keep the probability of enrollment unchanged, an increase in the probability of the job to be outside of state 0 to 100% should be compensated by a salary increase of Rs 5,000. Since the average salary is Rs 10,000, this suggests that prospective trainees expect a 50% higher salary for jobs outside of state. In the data we collected for a companion project (Chakravorty et al., 2024), DDU-GKY placement jobs out of state are only paid 3% higher than jobs in the state. Hence the results suggest that location preferences are a substantial barrier to program enrollment.

Since our experiment only manipulates jobs characteristics, keeping other aspects of the vocational training constant, our design allows us to identify preferences about jobs separately from preferences about training. Hence we can reasonably expect that the role of placement jobs' location and salary in enrollment decisions is similar to the trade-off involved in the actual migration decision.[18] This allows us to interpret our estimate of the salary premium for jobs out of state as a migration cost, and to benchmark our results against migration costs found in the literature. Our estimates are lower than Tombe and Zhu (2019)'s finding that inter-province migration costs in China are twice as large as within-province (0.97 vs 0.45). Using data from Chakravorty et al. (2024), we estimate that jobs out of state are located on average 10 times further away than jobs in the state. This implies an elasticity of migration costs to distance of 0.05, which is between Bryan and Morten (2019)'s estimate for Indonesia (0.15) and their estimate for the US (0.02). Hence our estimate of migration cost is on the lower side of the estimates available in the literature in other countries, which could partly be due to the fact that we focus on a sub-population of young job seekers. These migration costs are still large enough that they would give up good placement opportunities.

Moreover, the required 50% salary premium may not fully represent a net gain, as it overlooks additional frictions and non-monetary costs influencing participation decisions. First, the opportunity cost of training is substantial, as participants must forgo earnings from informal or agricultural work during the residential program. Even if these forgone earnings are low, they may weigh heavily on job seekers, particularly those contributing to household income. Second, uncertainty about placement jobs further reduces the perceived benefits of training. As Chakravorty et al. (2024) show, dropout rates remain high even after placement, suggesting that many trainees reconsider urban employment despite guaranteed jobs. Finally, non-monetary barriers, such as social and cultural constraints,

[18] The migration decision is usually made by DDU-GKY trainees at the end of the training but we can not observe it for our sample due to the closure of training centers when the COVID epidemic started.

particularly affect young workers from close-knit rural communities, further discouraging enrollment (Imbert and Papp, 2020). Together, these factors reinforce the idea that high expected migration costs—both monetary and non-monetary—act as a major obstacle to program participation.

Following our pre-analysis plan, we also explore heterogeneity by gender, caste, and education levels. We find evidence of higher migration costs for SC & ST than for higher castes (ratio of -3.9 and -1.3 respectively), and for less educated than more educated (ratios of -1.9 and 0.8 respectively), but the 95% confidence intervals overlap (Appendix Table A11).

5 Conclusion

This paper studies the role of information frictions and location preferences in the decision made by young rural workers to engage in a government-sponsored training scheme that guarantees placement into a formal job in urban areas, which may be located inside or outside their home state. We find that candidates were over-optimistic: they expected placement jobs to be closer to home and to pay more than they did. We experimentally informed them about the probability that jobs were outside the state, about the wage distribution, or both. We show that the intervention makes job seekers more pessimistic and changes their decision to enroll. Revealed preference estimates imply that rural job seekers require 50% higher pay to work outside of their home state. This suggests that large wage differentials are needed to compensate for the disutility of cross-state migration in India.

Our findings have broader implications for the design of government training programs. While we were unable to track long-term employment outcomes due to disruptions caused by the COVID-19 pandemic, our results highlight the role of misinformation at the mobilization stage. The DDU-GKY program includes mechanisms aimed at addressing this issue, such as extensive counseling sessions for candidates and their guardians before enrollment, which provide more detailed information about vocational training benefits and actual labor market prospects. However, our results suggest that these later-stage counseling efforts may not fully correct misperceptions.

A simple information intervention at the mobilization stage, before individuals reach the training center, can significantly alter labor market beliefs and influence enrollment decisions. Future research could explore the relative effectiveness of providing information at different stages—early mobilization versus later counseling—in shaping labor market expectations and long-term job retention. Additionally, alternative incentive structures for training providers could be investigated, particularly those that align their incentives not just with initial enrollment but also with training completion and job retention. Understanding how best to structure these programs to ensure both informed enrollment decisions and long-term job retention remains an important avenue for future research.

References

Abebe, G., A. S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2021a): "Anonymity or distance? Job search and labour market exclusion in a growing African city," *The Review of Economic Studies*, 88, 1279–1310.

Abebe, G., A. S. Caria, and E. Ortiz-Ospina (2021b): "The selection of talent: Experimental and structural evidence from ethiopia," *American Economic Review*, 111, 1757–1806.

Abebe, G. T., S. Caria, M. Fafchamps, P. Falco, S. Franklin, S. Quinn, and F. Shilpi (2017): "Job fairs: Matching firms and workers in a field experiment in Ethiopia," *World Bank Policy Research Working Paper*.

Adhvaryu, A., N. Kala, and A. Nyshadham (2023): "Returns to on-the-job soft skills training," *Journal of Political Economy (forthcoming)*.

Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2020): "Tackling youth unemployment: Evidence from a labor market experiment in Uganda," *Econometrica*, 88, 2369–2414.

Alfonsi, L., M. Namubiru, and S. Spaziani (2022): "Gender Gaps: Back and Here to Stay? Evidence from Skilled Ugandan Workers during COVID-19," .

Altmann, S., A. Falk, S. Jäger, and F. Zimmermann (2018): "Learning about job search: A field experiment with job seekers in Germany," *Journal of Public Economics*, 164, 33–49.

Bah, T. L. and C. Batista (2018): "Understanding willingness to migrate illegally: Evidence from a lab in the field experiment," *Nova School of Business and Economics Working Paper*.

Bah, T. L., C. Batista, F. Gubert, and D. McKenzie (2022): "Can information and alternatives to irregular migration reduce "backway" migration from the Gambia?" NOVAFRICA Working Paper Series wp2203.

Bandiera, O., V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2023): "The Search for Good Jobs: Evidence from a Six Year Field Experiment in Uganda," Manuscript.

Banerjee, A. and G. Chiplunkar (2022): "How important are matching frictions in the labor market? experimental & non-experimental evidence from a large Indian firm," *Working Paper*.

Banerjee, A. V. and S. Sequeira (2020): "Spatial mismatches and imperfect information in the job search," *CEPR Discussion Paper No. DP14414*.

Baseler, T. (2022): "Hidden income and the perceived returns to migration," *American Economic Journal: Applied Economics (forthcoming)*.

Bassi, V. and A. Nansamba (2022): "Screening and signalling non-cognitive skills: experimental evidence from Uganda," *The Economic Journal*, 132, 471–511.

Batista, C. and D. J. McKenzie (2021): "Testing classic theories of migration in the lab," *CEPR Discussion Paper No. DP16469*.

Bazzi, S., L. Cameron, S. G. Schaner, and F. Witoelar (2021): "Information, intermediaries, and international migration," Tech. rep., National Bureau of Economic Research.

Bell, M., E. Charles-Edwards, P. Ueffing, J. Stillwell, M. Kupiszewski, and D. Kupiszewska (2015): "Internal migration and development: Comparing migration intensities around the world," *Population and Development Review*, 41, 33–58.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81, 608–650.

Belot, M., P. Kircher, and P. Muller (2019): "Providing advice to jobseekers at low cost: An experimental study on online advice," *The Review of Economic Studies*, 86, 1411–1447.

——— (2021): "Eliciting time preferences when income and consumption vary: Theory, validation & application to job search," *Tinbergen Institute Discussion Paper 2021-013/V*.

Blattman, C. and S. Dercon (2018): "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 10, 1–38.

Blattman, C. and L. Ralston (2015): "Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs," *Available at SSRN 2622220*.

Bloom, N., G. Fischer, I. Rasul, A. Rodriguez-Clare, T. Suri, C. Udry, E. Verhoogen, C. Woodruff, and G. Zane (2014): "Firm Capabilities and Economic Growth," *Evidence Paper, International Growth Centre. Disponível em: https://www. theigc. org/wp-content/uploads/2014/09/IGCEvidencePaperFirms. pdf*.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014): "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh," *Econometrica*, 82, 1671–1748.

Bryan, G. and M. Morten (2019): "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 127, 2229–2268.

Carranza, E., R. Garlick, K. Orkin, and N. Rankin (2020): "Job Search and Hiring With Two-Sided Limited Information About Workseekers' Skills," *Economic Research Initiatives at Duke (ERID) Working Paper*.

Chakravorty, B., W. Arulampalam, A. Y. Bhatiya, C. Imbert, and R. Rathelot (2024): "Can information about jobs improve the effectiveness of vocational training? Experimental evidence from India," *Journal of Development Economics*, 169, 103273.

Chakravorty, B., A. Y. Bhatiya, C. Imbert, M. Lohnert, P. Panda, and R. Rathelot (2023): "Impact of the COVID-19 crisis on India's rural youth: Evidence from a panel survey and an experiment," *World Development*, 168, 106242.

Correia, S. (2015): "Singletons, cluster-robust standard errors and fixed effects: A bad mix," *Technical Note, Duke University*, 7.

Cullen, Z. and R. Perez-Truglia (2022): "How much does your boss make? The effects of salary comparisons," *Journal of Political Economy*, 130, 766–822.

Delavande, A., X. Giné, and D. McKenzie (2011): "Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design?" *Journal of Applied Econometrics*, 26, 479–497.

Franklin, S. (2018): "Location, search costs and youth unemployment: experimental evidence from transport subsidies," *The Economic Journal*, 128, 2353–2379.

Frohnweiler, S., B. Beber, and C. Ebert (2022): "Information Frictions, Belief Updating and Internal Migration: Evidence from Ghana and Uganda," *IZA Discussion Paper*.

Fuster, A. and B. Zafar (2023): "Survey experiments on economic expectations," in *Handbook of Economic Expectations*, Elsevier, 107–130.

Gollin, D., D. Lagakos, and M. E. Waugh (2014): "The agricultural productivity gap," *The Quarterly Journal of Economics*, 129, 939–993.

Hicks, J. H., M. Kremer, I. Mbiti, and E. Miguel (2011): "Vocational education voucher delivery and labor market returns: A randomized evaluation among Kenyan youth," *Report for Spanish Impact Evaluation Fund (SIEF) Phase II*.

Imbert, C. and J. Papp (2020): "Costs and benefits of rural-urban migration: Evidence from India," *Journal of Development Economics*, 146, 102473.

Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2022): "Worker beliefs about outside options," *The Quarterly Journal of Economics (forthcoming)*.

Jensen, R. (2012): "Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India," *The Quarterly Journal of Economics*, 127, 753–792.

Jones, S. and R. Santos (2022): "Can information correct optimistic wage expectations? Evidence from Mozambican job-seekers," *Journal of Development Economics*, 159, 102987.

Kircher, P. (2022): "Job search in the 21st Century," *Journal of the European Economic Association*, 20, 2317–2352.

Kone, Z. L., M. Y. Liu, A. Mattoo, C. Ozden, and S. Sharma (2018): "Internal borders and migration in India," *Journal of Economic Geography*, 18, 729–759.

Lagakos, D., A. M. Mobarak, and M. E. Waugh (2023): "The welfare effects of encouraging rural-urban migration," *Econometrica*, 91, 803–837.

Manning, A. and B. Petrongolo (2017): "How Local Are Labor Markets? Evidence from a Spatial Job Search Model," *American Economic Review*, 107, 2877–2907.

Marinescu, I. and R. Rathelot (2018): "Mismatch unemployment and the geography of job search," *American Economic Journal: Macroeconomics*, 10, 42–70.

Mas, A. and A. Pallais (2017): "Valuing Alternative Work Arrangements," *American Economic Review*, 107, 3722–3759.

McKenzie, D. (2017): "How effective are active labor market policies in developing countries? a critical review of recent evidence," *The World Bank Research Observer*, 32, 127–154.

McKenzie, D., J. Gibson, and S. Stillman (2013): "A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad?" *Journal of Development Economics*, 102, 116–127.

McKenzie, D. and D. Yang (2022): "Field and Natural Experiments in Migration," *World Bank, Washington, DC*.

Meghir, C., A. M. Mobarak, C. Mommaerts, and M. Morten (2022): "Migration and informal insurance: evidence from a randomized controlled trial and a structural model," *The Review of Economic Studies*, 89, 452–480.

Munshi, K. and M. Rosenzweig (2016): "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 106, 46–98.

Schmutz, B. and M. Sidibé (2019): "Frictional Labour Mobility," *Review of Economic Studies*, 86, 1779–1826.

Shrestha, M. (2020): "Get Rich or Die Tryin': Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal," *The World Bank Economic Review*, 34, 1–27.

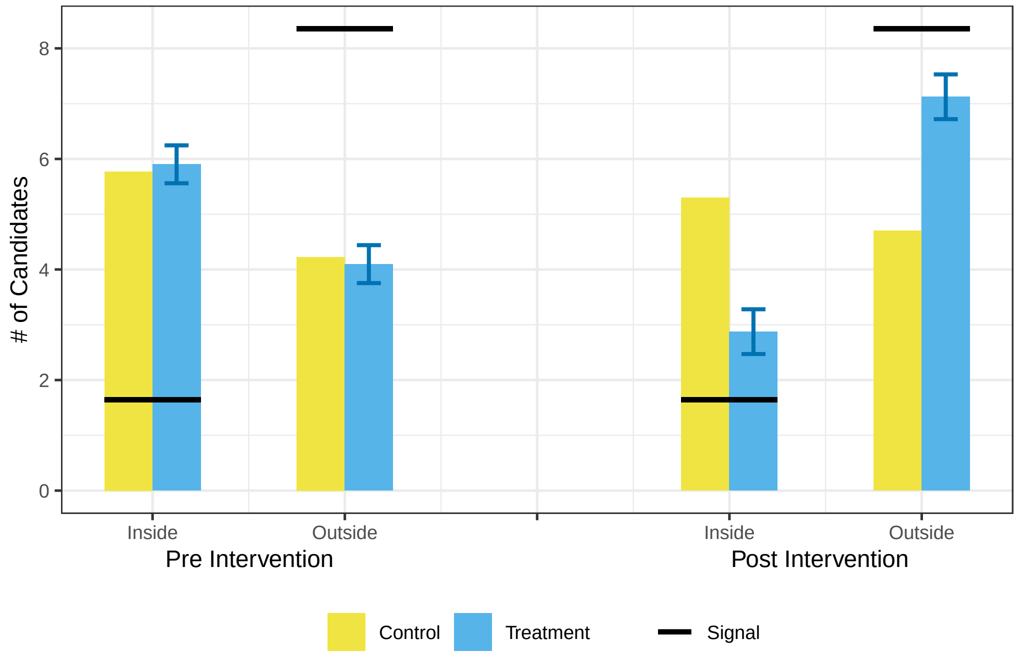
Tombe, T. and X. Zhu (2019): "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 109, 1843–72.

Van Ommeren, J. and M. Fosgerau (2009): "Workers' marginal costs of commuting," *Journal of Urban Economics*, 65, 38–47.

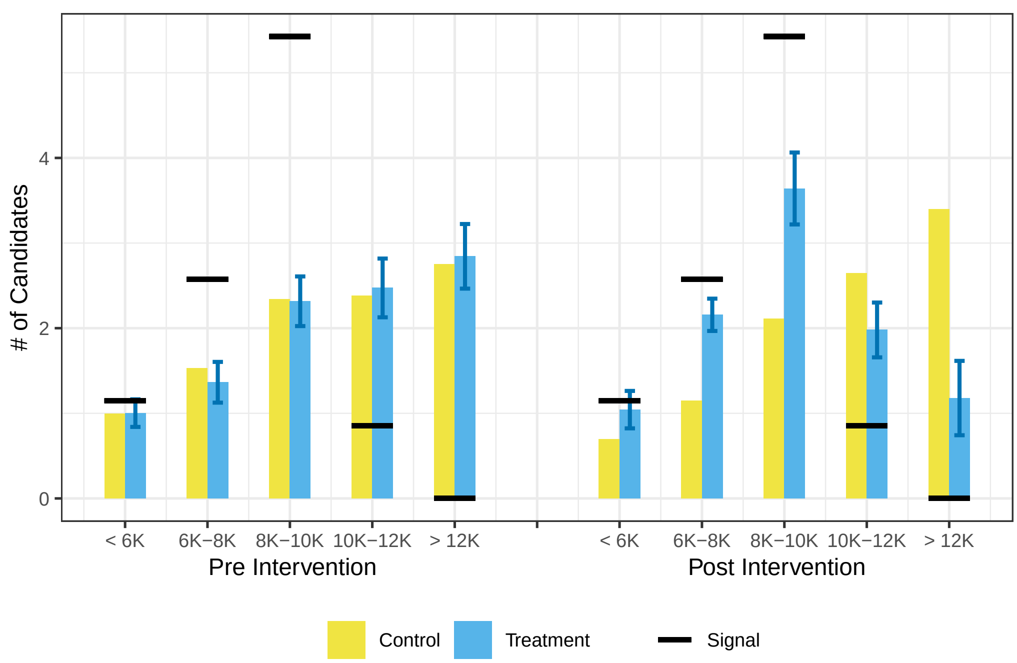
Young, A. (2013): "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics*, 128, 1727–1785.

Tables and Figures

Figure 1: Effect of Treatment on Labor Market Beliefs



(a) Location



(b) Salary

Notes: Panel A shows the distribution of the number of candidates inside and outside the state from the baseline survey pre- and post-intervention. Panel B shows the distribution of number of candidates from the baseline survey pre- and post-intervention in 5 salary bins: below Rs 6K, 6K-8K, 8K-10K, 10K-12K and above Rs 12K. The error bars show the 95% CI on the coefficient of an indicator variable for the information treatment. "Signal" is the information provided by the treatment videos.

Table 1: Effect of Treatment on Labor Market Beliefs

	Posterior − Prior		
	Baseline Posterior	Followup 1w	Followup 4w
	(1)	(2)	(3)
Panel A: Location (Candidates Outside State)			
Location Treatment	2.495***	1.276***	1.228***
	(0.223)	(0.221)	(0.254)
Mean DV [Control]	0.474	0.890	0.812
Prior [Control]	4.227	4.291	4.215
Panel B: Salary (Earnings Distribution Mean)			
Salary Treatment	-1.463***	-0.655***	-0.633***
	(0.125)	(0.132)	(0.129)
Mean DV [Control]	0.506	0.001	0.117
Prior [Control]	9.873	9.856	9.886
# of Camps	63	62	63
Camp FE	Yes	Yes	Yes
Observations	876	823	826

Notes: This table presents the effect of the location treatment (Panel A) and the salary treatment (Panel B) on how the respondents update their labor market beliefs (Posterior - Prior). Column 1 measures the outcomes after the intervention during the baseline survey. Columns 2 and 3 measure the outcomes at the follow-up surveys one week and four weeks after the intervention, respectively. The outcome variables in Panel A measure the number of candidates (out of 10) who will get a job outside state. The outcome variables in Panel B measure earnings distribution mean calculated using the number of candidates in each bin. All outcomes in Panel B are scaled by 1000. Standard errors are clustered at the camp level. The control variables chosen by a post-double-selection lasso procedure were a dummy for females and having the RSBY document (Panel A) and a dummy variable for having the NREGA job card (Panel B). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Effect of Treatment on Own Career Expectations (1 year later)

	Posterior		
	(1) Baseline Posterior	(2) Followup 1w	(3) Followup 4w
Panel A: Respondent Outside of State if Completes Training			
Location Treatment	0.072**	0.096***	0.097***
	(0.028)	(0.032)	(0.032)
Mean DV [Control]	0.337	0.396	0.376
Panel B: Respondent Outside of State if Does Not Complete Training			
Location Treatment	-0.010	-0.016	0.013
	(0.018)	(0.020)	(0.019)
Mean DV [Control]	0.113	0.093	0.077
Panel C: Respondent Salary if Completes Training			
Salary Treatment	-1.700***	-1.049***	-1.094***
	(0.299)	(0.316)	(0.299)
Mean DV [Control]	13.173	13.959	13.442
Panel D: Respondent Salary if Does Not Complete Training			
Salary Treatment	-0.440	-0.133	0.093
	(0.357)	(0.431)	(0.420)
Mean DV [Control]	6.358	7.470	7.128
# of Camps	63	62	63
Camp FE	Yes	Yes	Yes
Observations	876	823	825

Notes: This table presents the estimation of own career expectations (location and salary) one year later on the treatment status. Column 1 measures the outcomes after the intervention during the baseline survey. Columns 2 and 3 measure outcomes at the follow-up surveys one week and four weeks after the intervention, respectively. The dependent variable in Panels A and B indicates whether the respondent expects to be outside of Bihar one year later with training (expectation; Panel A) and without training (counterfactual; Panel B). The dependent variables in Panels C and D measure the average monthly salary one year later with training (expectation; Panel C) and without training (counterfactual; Panel D). All outcomes in Panels C and D are scaled by 1000. Standard errors are clustered at the camp level. The control variables were chosen by a post-double-selection lasso procedure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Heterogeneity by Signal (First Stage Regressions)

	Posterior − Prior			
	Location		Salary	
	(1) Followup 1w	(2) Followup 4w	(3) Followup 1w	(4) Followup 4w
Location Treatment	0.439*	0.261	0.137	0.398**
	(0.251)	(0.324)	(0.147)	(0.171)
Location (Signal − Prior)	0.653***	0.595***	0.083***	0.108***
	(0.060)	(0.052)	(0.025)	(0.027)
Location (Signal − Prior) × Treatment Location	0.215***	0.236***	-0.044	-0.080**
	(0.051)	(0.063)	(0.032)	(0.039)
Salary Treatment	0.133	0.145	-0.396***	-0.258*
	(0.289)	(0.223)	(0.118)	(0.147)
Salary (Signal − Prior)	0.027	0.113	0.655***	0.636***
	(0.107)	(0.079)	(0.044)	(0.048)
Salary (Signal − Prior) × Salary Treatment	0.211*	0.110	0.137**	0.182***
	(0.124)	(0.104)	(0.055)	(0.065)
Mean DV [Control]	0.890	0.812	0.001	0.117
Prior [Control]	4.291	4.215	9.856	9.886
# of Camps	62	63	62	63
Camp FE	Yes	Yes	Yes	Yes
Observations	823	826	823	826

Notes: This table presents the first stage regressions as described in Section 3. The dependent variables in Columns 1 and 2 measure the number of candidates (out of 10) who will get a job outside state. The dependent variables in Columns 3 and 4 measure earnings distribution mean calculated using the number of candidates in each bin. All outcomes in Columns 3 and 4 are scaled by 1000. Followup 1w and Folloupw 4w indicate the follow-up surveys one week and four weeks after the intervention respectively. Standard errors are clustered at the camp level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Effect of Beliefs on Training Enrollment

	I(Enrollment) − P(Enrollment Prior)		
	(1) Followup 1w	(2) Followup 4w	(3) Admin
Location (Posterior − Prior)	-0.007	-0.010	-0.012**
	(0.007)	(0.006)	(0.006)
Salary (Posterior − Prior)	0.021**	0.024**	0.024***
	(0.010)	(0.011)	(0.008)
Mean DV [Control]	-0.602	-0.544	-0.684
P(Enrollment Prior) [Control]	0.787	0.787	0.787
Enrollment [Control]	0.187	0.238	0.103
KP F Stat	67.59	62.22	62.22
F Stat (Salary)	130.7	145.3	145.3
F Stat (Location)	70.54	112.37	112.37
Bootstrapped Ratio Mean	-2.81	-2.55	-2.14
Bootstrapped Ratio 95% CI	[-21.76, 10.99]	[-19.17, 17.40]	[-5.77, -0.50]
# of Camps	62	63	63
Camp FE	Yes	Yes	Yes
Observations	823	826	826

Notes: This table presents the 2SLS estimates of the respondents' updated beliefs on their updated probability to enroll in the training program. The updated beliefs are instrumented using the treatment status as described in Section 3. Columns 1 and 2 measure outcomes at the follow-up surveys one week and four weeks after the intervention respectively. Column 3 measures the enrollment outcome from the administrative dataset. The KP F stat is the Kleibergen-Paap F-stat for the joint significance of the two instruments in the first-stage regression. The F-stat (Salary) and F-stat (Location) are the Sanderson-Windmeijer partial F-stat for the instruments' joint significance in the two separate first-stage regressions. Standard errors are clustered at the camp level. The control variables chosen by a post-double-selection lasso procedure were a dummy for having the NREGA job card. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix

A Appendix Tables and Figures

Table A1: Summary Statistics

Variable	Mean	SD	Minimum	Maximum	Signal
Panel A: Socio-Demographics (Full Sample)					
Female	0.575	0.495	0	1	
Age	20.42	3.366	17	35	
I(Education $\geq$ Higher Secondary)	0.578	0.494	0	1	
Religion: Hindu	0.929	0.257	0	1	
Social Category: SC or ST	0.303	0.460	0	1	
Social Category: OBC	0.556	0.497	0	1	
Social Category: General	0.122	0.328	0	1	
Social Category: Prefer No Answer	0.0194	0.138	0	1	
Number of Observations			876		
Panel B: Prior Labor Market Beliefs (Females)					
Location (Candidates Outside State)	3.762	2.774	0	10	9
Salary (monthly average - Rs)	9836	1679	5000	13000	7600
... Less than Rs 6000 per month	1.050	1.542	0	10	2
... Rs 6000 - Rs 8000 per month	1.597	1.814	0	10	3
... Rs 8000 - Rs 10,000 per month	2.179	2.308	0	10	5
... Rs 10,000 - Rs 12,000 per month	2.472	2.528	0	10	0
... More than Rs 12,000 per month	2.702	2.967	0	10	0
Difficulty to family during training [0-10]	3.014	3.642	0	10	
P(Enrollment)	0.802	0.277	0	1	
Panel C: Prior Labor Market Beliefs (Males)					
Location (Candidates Outside State)	5.215	2.557	0	10	7
Salary (monthly average - Rs)	9824	1708	5000	13000	9000
... Less than Rs 6000 per month	1.097	1.622	0	10	0
... Rs 6000 - Rs 8000 per month	1.527	1.632	0	10	2
... Rs 8000 - Rs 10,000 per month	2.202	1.798	0	10	6
... Rs 10,000 - Rs 12,000 per month	2.511	2.219	0	10	2
... More than Rs 12,000 per month	2.664	2.825	0	10	0
Difficulty to family during training [0-10]	3.618	3.439	0	10	
P(Enrollment)	0.767	0.286	0	1	

Notes: This table presents summary statistics on socio-demographic characteristics (Panel A) for the full sample and prior labor market beliefs for females (Panel B) and males (Panel C). The prior labor market beliefs are presented separately by gender due to differences in signal by gender.

Table A2: Balance Statistics

Variable	Control Mean	Treatment Mean			p-value		
		Salary	Location	Salary × Location			
	(1)	(2)	(3)	(4)	(2) vs (1)	(3) vs (1)	(4) vs (1)
Panel A: Socio-Demographic Variables							
Female	0.580	0.550	0.547	0.624	0.527	0.494	0.361
Age	20.31	20.47	20.42	20.47	0.624	0.749	0.626
I(Education $\geq$ Higher Secondary)	0.591	0.558	0.561	0.603	0.504	0.536	0.805
Religion: Hindu	0.927	0.939	0.924	0.926	0.634	0.884	0.946
Religion: Muslim	0.0466	0.0519	0.0314	0.0437	0.789	0.448	0.882
Religion: Prefer No Answer	0.0259	0.00866	0.0448	0.0306	0.278	0.238	0.770
Social Category: SC or ST	0.290	0.303	0.296	0.319	0.774	0.898	0.525
Social Category: OBC	0.591	0.550	0.534	0.555	0.400	0.244	0.458
Social Category: General	0.0933	0.121	0.157	0.114	0.382	0.0482	0.526
Social Category: Prefer No Answer	0.0259	0.0260	0.0135	0.0131	0.996	0.359	0.343
Panel B: Prior Labor Market Beliefs							
Location (Candidates Outside State)	4.249	4.641	4.345	4.258	0.148	0.724	0.974
Salary (monthly average - Rs)	9860	9791	9684	9989	0.677	0.290	0.436
... Less than Rs 6000 per month	1	1.087	1.139	1.044	0.574	0.370	0.777
... Rs 6000 - Rs 8000 per month	1.539	1.649	1.749	1.332	0.514	0.218	0.222
... Rs 8000 - Rs 10,000 per month	2.352	2.100	2.148	2.179	0.219	0.324	0.400
... Rs 10,000 - Rs 12,000 per month	2.378	2.550	2.480	2.528	0.464	0.667	0.523
... More than Rs 12,000 per month	2.731	2.615	2.484	2.917	0.683	0.389	0.512
Difficulty to family during training [0-10]	3.352	2.861	3.552	3.341	0.158	0.570	0.973
Difficulty to family 1 year outside State [0-10]	4.021	3.450	3.583	3.812	0.120	0.237	0.571
P(Enrollment)	0.786	0.786	0.792	0.783	0.996	0.831	0.925
Number of Observations				880			

Notes: To check that our randomization achieved a balance between treatment and control at baseline, we estimate the following model for each control variable X_i^l and test the null of no difference between the treatment groups and control group ($\beta_s = 0$, $\beta_l = 0$ and $\beta_{sl} = 0$).

$$X_i = \beta_s T_i^s + \beta_l T_i^l + \beta_{sl} T_i^s \times T_i^l + \varepsilon_i$$

Table A3: Effect of Camp Characteristics on Misperceptions in Beliefs

	Misperceptions in Prior Beliefs	
	(1) Location	(2) Salary
I(Mobilizer in Camp)	0.023	-0.009
	(0.050)	(0.031)
I(District: Muzaffarpur)	-0.040	0.012
	(0.092)	(0.061)
I(District: Nawada)	0.065	0.052
	(0.090)	(0.063)
I(District: Samastipur)	0.055	0.044
	(0.078)	(0.060)
Observations	770	770

Notes: This table presents the effect of the camp characteristics on misperceptions in labor market beliefs. Misperceptions are calculated as the percentage difference in the prior and signal. We use dummy variables for presence of a mobilizer in the camp and the district of the camp. Standard errors are clustered at the camp level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Effect of Socio-Demographic Characteristics on Misperceptions in Beliefs

	Misperceptions in Prior Beliefs			
	(1) Location	(2) Location	(3) Salary	(4) Salary
I(Female)	-0.289***	-0.294***	0.223***	0.210***
	(0.032)	(0.039)	(0.018)	(0.019)
log(Age)	-0.008	0.026	-0.066	-0.019
	(0.074)	(0.073)	(0.044)	(0.047)
I(Education $\geq$ Higher Secondary)	0.044*	0.041*	0.016	0.011
	(0.024)	(0.022)	(0.017)	(0.018)
Religion: Muslim	-0.055	0.010	-0.048	-0.048
	(0.036)	(0.047)	(0.038)	(0.049)
Social Category: SC or ST	-0.028	-0.036	-0.005	-0.005
	(0.031)	(0.035)	(0.023)	(0.026)
Social Category: OBC	-0.059**	-0.085**	0.025	0.026
	(0.028)	(0.034)	(0.023)	(0.027)
Camp FE	No	Yes	No	Yes
Observations	880	876	880	876

Notes: This table presents the effect of the socio-demographic characteristics on misperceptions in labor market beliefs. Misperceptions are calculated as the percentage difference in the prior and signal. For Location, we measure the number of candidates (out of 10) who will get a job outside state. On Salary, we measure earnings distribution mean calculated using the number of candidates in each bin. Standard errors are clustered at the camp level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Attrition

	Attrition	
	(1) Followup 1w	(2) Followup 4w
Location Treatment	-0.001	-0.012
	(0.025)	(0.028)
Salary Treatment	-0.011	-0.026
	(0.024)	(0.025)
Location Treatment $\times$ Salary Treatment	0.006	0.008
	(0.032)	(0.033)
Mean DV [Control]	0.062	0.067
# of Camps	63	63
Camp FE	Yes	Yes
Observations	876	876

Notes: Standard errors are clustered at the camp level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Effect of Treatment on Salary Distribution: Mean, Median and Variance

	Posterior − Prior		
	(1) Baseline Posterior	(2) Followup 1w	(3) Followup 4w
Panel A: Mean of Salary Distribution			
Salary Treatment	-1.463***	-0.655***	-0.633***
	(0.125)	(0.132)	(0.129)
Mean DV [Control]	0.506	0.001	0.117
Prior [Control]	9.873	9.856	9.886
Panel B: Median of Salary Distribution			
Salary Treatment	-1.643***	-0.740***	-0.655***
	(0.159)	(0.170)	(0.169)
Mean DV [Control]	0.753	0.187	0.155
Prior [Control]	9.567	9.560	9.608
Panel C: Variance of Salary Distribution			
Salary Treatment	-1.221***	-0.464	-0.787*
	(0.312)	(0.408)	(0.453)
Mean DV [Control]	0.289	0.899	1.235
Prior [Control]	5.670	5.752	5.657
# of Camps	63	62	63
Camp FE	Yes	Yes	Yes
Observations	876	823	826

Notes: This table presents the effect of the salary treatment on the changes in the mean (Panel A), median (Panel B) and the variance (Panel C) of the salary distribution. Column 1 measures the outcomes after the intervention during the baseline survey. Columns 2 and 3 measure the outcomes at the followup survey one week and four weeks after the intervention respectively. All outcomes in Panel A are scaled by 1000. Standard errors are clustered at the camp level. Control variables were selected using a post double-selection lasso. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Effect of Treatment on Labor Market Beliefs

	Posterior − Prior		
	Baseline Posterior	Followup 1w	Followup 4w
	(1)	(2)	(3)
Panel A: Location (Candidates Outside State)			
Location Treatment	2.433***	1.342***	1.235***
	(0.316)	(0.291)	(0.327)
Salary Treatment	-0.402	-0.240	-0.245
	(0.284)	(0.309)	(0.272)
Location Treatment × Salary Treatment	0.095	-0.142	-0.030
	(0.436)	(0.408)	(0.427)
Mean DV [Control]	0.474	0.890	0.812
Prior [Control]	4.227	4.291	4.215
Panel B: Salary (Earnings Distribution Mean)			
Location Treatment	-0.059	0.055	0.080
	(0.153)	(0.143)	(0.189)
Salary Treatment	-1.380***	-0.442**	-0.509***
	(0.153)	(0.181)	(0.174)
Location Treatment × Salary Treatment	-0.168	-0.417*	-0.241
	(0.223)	(0.234)	(0.253)
Mean DV [Control]	0.506	0.001	0.117
Prior [Control]	9.873	9.856	9.886
# of Camps	63	62	63
Camp FE	Yes	Yes	Yes
Observations	876	823	826

Notes: This table presents the effect of the location treatment (Panel A) and the salary treatment (Panel B) on how the respondents update their labor market beliefs (Posterior - Prior). Column 1 measures the outcomes after the intervention during the baseline survey. Columns 2 and 3 measure the outcomes at the follow-up surveys one week and four weeks after the intervention, respectively. The outcome variables in Panel A measure the number of candidates (out of 10) who will get a job outside state. The outcome variables in Panel B measure earnings distribution mean calculated using the number of candidates in each bin. All outcomes in Panel B are scaled by 1000. Standard errors are clustered at the camp level. The control variables chosen by a post-double-selection lasso procedure were a dummy for females and having the RSBY document (Panel A) and a dummy variable for having the NREGA job card (Panel B). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Spillover Effects of Treatment

	Posterior − Baseline Posterior			
	Followup 1w		Followup 4w	
	(1)	(2)	(3)	(4)
Panel A: Location (Candidates Outside State)				
Location Treatment	-1.128***	-2.336***	-1.376***	-1.972***
	(0.230)	(0.643)	(0.219)	(0.669)
Share Treated		-1.803***		0.117
		(0.671)		(0.594)
Location Treatment × Share Treated		2.622**		0.853
		(1.071)		(0.980)
Mean DV [Control]	0.407	0.407	0.464	0.464
Baseline Posterior [Control]	4.701	4.701	4.701	4.701
p-value: Share+Interaction		0.279		0.240
Panel B: Salary (Earnings Distribution Mean)				
Salary Treatment	0.755***	-1.003**	0.793***	-0.665*
	(0.106)	(0.404)	(0.122)	(0.380)
Share Treated		-0.307		-0.178
		(0.455)		(0.415)
Salary Treatment × Share Treated		0.656		0.122
		(0.698)		(0.675)
Mean DV [Control]	-0.520	-0.520	-0.385	-0.385
Baseline Posterior [Control]	10.379	10.379	10.379	10.379
p-value: Share+Interaction		0.461		0.905
Camp FE	Yes	Yes	Yes	Yes
Observations	823	823	826	826

Notes: This table presents the effect of the treatment on labor market beliefs for location (Panel A) and salary (Panel B). The share of treated respondents is defined as the share of treatment within a peer group (defined as mobilization camp × panchayat). The outcome variables in Panel A measure the number of candidates (out of 10) who will get a job outside state. The outcome variables in Panel B measure earnings distribution mean calculated using the number of candidates in each bin. All outcomes in Panel B are scaled by 1000. Standard errors are clustered at the camp level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Effect of Beliefs on Training Enrollment Intentions

	Probability to Enroll (Posterior − Prior)		
	(1) Baseline Posterior	(2) Followup 1w	(3) Followup 4w
Location (Posterior − Prior)	0.002	-0.003	-0.009
	(0.002)	(0.006)	(0.007)
Salary (Posterior − Prior)	0.006	0.010	0.013
	(0.005)	(0.008)	(0.010)
Mean DV [Control]	0.049	-0.123	-0.175
P(Enrollment) [Control]	0.787	0.787	0.787
KP F Stat	99.24	67.59	62.22
F Stat (Salary)	123.7	130.7	145.3
F Stat (Location)	102.1	70.54	112.37
# of Camps	63	62	63
Camp FE	Yes	Yes	Yes
Observations	876	823	826

Notes: This table presents the 2SLS estimates of the respondents' updated beliefs on their updated probability to enroll the training program. The updated beliefs are instrumented using the treatment status as described in the Section 3. Column 1 measures the outcomes after the intervention during the baseline survey. Columns 2 and 3 measure outcomes at the follow-up surveys one week and four weeks after the intervention respectively. The KP F stat is the Kleibergen-Paap F-stat for the joint significance of the two instruments in the first-stage regression. The F-stat (Salary) and F-stat (Location) are the Sanderson-Windmeijer partial F-stat for the instruments' joint significance in the two separate first-stage regressions. Standard errors are clustered at the camp level. Control variables selected using post double-selection lasso. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Effect of Treatment on Training Enrollment

	I(Enrollment) − P(Enrollment Prior)		
	(1)	(2)	(3)
	Followup 1w	Followup 4w	Admin
Panel A: Any Treatment			
Any Treatment	-0.045	-0.062*	-0.039
	(0.037)	(0.036)	(0.030)
Panel B: All Treatment Arms			
Location Treatment	-0.049	-0.027	-0.046
	(0.046)	(0.047)	(0.039)
Salary Treatment	-0.051	-0.097**	-0.035
	(0.045)	(0.038)	(0.032)
Location Treatment ×	0.063	0.067	0.046
Salary Treatment	(0.059)	(0.061)	(0.043)
Mean DV [Control]	-0.602	-0.544	-0.684
Prior [Control]	0.787	0.787	0.787
Admission [Control]	0.187	0.238	0.103
# of Camps	62	63	63
Camp FE	Yes	Yes	Yes
Observations	823	826	826

Notes: This table presents the reduced form estimates of the treatment on the updated probability to enroll in the training program. In Panel A, Any Treatment is a dummy variable that takes a value of one for respondents receiving either treatment and zero for the control group. In Panel B, we consider all treatment arms. Columns 1 and 2 measure outcomes at the follow-up surveys one week and four weeks after the intervention respectively. Column 3 measures the enrollment outcome from the administrative dataset. Standard errors are clustered at the camp level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A11: Effect of Beliefs on Training Enrollment
(Bootstrapped across sub-samples)

	I(Enrollment) − P(Enrollment Prior)						
	(1) Full Sample	(2) Females	(3) Males	(4) General & OBC	(5) SC & ST	(6) High Educ	(7) Low Educ
Location (Posterior − Prior)	-0.012**	-0.004	-0.011	-0.013*	-0.010	0.005	-0.029***
	(0.006)	(0.007)	(0.010)	(0.007)	(0.013)	(0.007)	(0.009)
Salary (Posterior − Prior)	0.024***	0.022	0.015	0.015	0.051**	0.009	0.051***
	(0.008)	(0.013)	(0.015)	(0.011)	(0.022)	(0.009)	(0.016)
Mean DV [Control]	-0.684	-0.734	-0.615	-0.707	-0.623	-0.678	-0.692
Prior [Control]	0.787	0.813	0.751	0.805	0.748	0.800	0.768
Enrollment [Control]	0.103	0.080	0.136	0.098	0.125	0.122	0.076
Bootstrapped Ratio Mean	-2.14	0.157	-0.841	-1.267	-3.896	0.780	-1.855
Bootstrapped Ratio 95% CI	[-5.77, -0.50]	[-24.07, 85.24]	[-17.70, 9.08]	[-9.38, 1.95]	[-103.62, 50.54]	[-20.51, 13.21]	[-4.58, -0.60]
# of Camps (Clusters)	63	52	51	58	45	58	53
Camp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	826	473	345	555	244	483	334

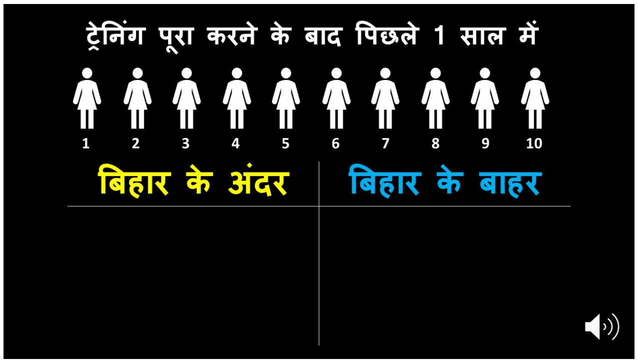
Notes: The total number of replications for each column is 200. Bootstrapped standard errors are clustered at the camp level. Control variables selected using post double-selection lasso. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A12: Spillover Effects of Treatment (not pre-specified)

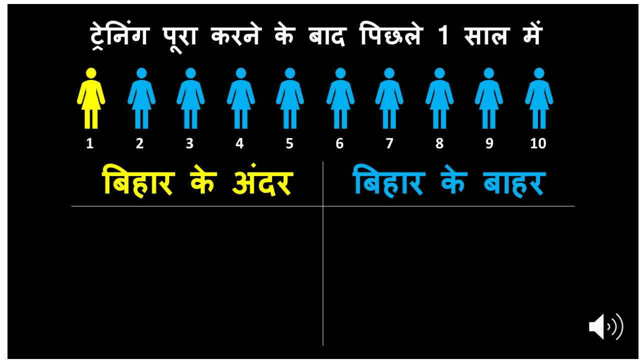
	Labor Market Beliefs			
	(1)	(2)	(3)	(4)
	Prior	Baseline Posterior	Followup 1w	Followup 4w
Panel A: Location (Candidates Outside Bihar)				
Location Treatment	-0.454	2.554***	0.191	0.661
	(0.524)	(0.405)	(0.534)	(0.514)
Share Treated	0.177	0.085	-1.818***	-0.016
	(0.498)	(0.486)	(0.605)	(0.542)
Location Treatment × Share Treated	0.417	-0.237	2.405**	0.776
	(0.859)	(0.641)	(0.905)	(0.826)
Mean DV [Control]	4.227	4.701	5.181	5.028
p-value: Share+Interaction	0.374	0.802	0.347	0.198
Panel B: Salary (Earnings Distribution Mean)				
Salary Treatment	0.291	-1.590***	-0.820***	-0.400
	(0.324)	(0.275)	(0.265)	(0.260)
Share Treated	0.594	0.079	0.271	0.384
	(0.447)	(0.437)	(0.344)	(0.431)
Salary Treatment × Share Treated	-0.560	0.317	0.232	-0.396
	(0.589)	(0.553)	(0.462)	(0.456)
Mean DV [Control]	9.873	10.379	9.857	10.003
p-value: Share+Interaction	0.929	0.266	0.095	0.969
Camp FE	Yes	Yes	Yes	Yes
Observations	876	876	823	826

Notes: This table presents the effect of the treatment on labor market beliefs for location (Panel A) and salary (Panel B). The share of treated respondents is defined as the share of treatment within a peer group (defined as mobilization camp × panchayat). The outcome variables in Panel A measure the number of candidates (out of 10) who will get a job outside state. The outcome variables in Panel B measure earnings distribution mean calculated using the number of candidates in each bin. All outcomes in Panel B are scaled by 1000. Standard errors are clustered at the camp level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

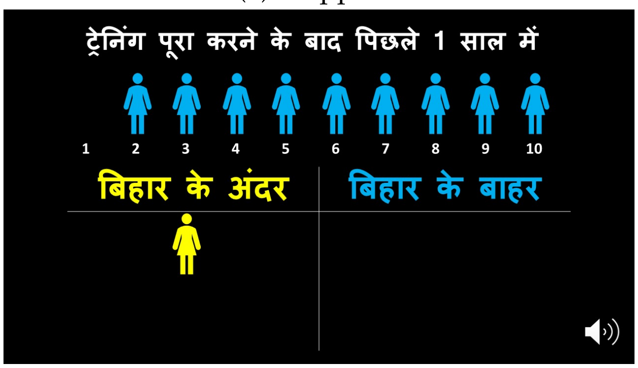
Figure A1: Location Intervention Video Snippets (Female)



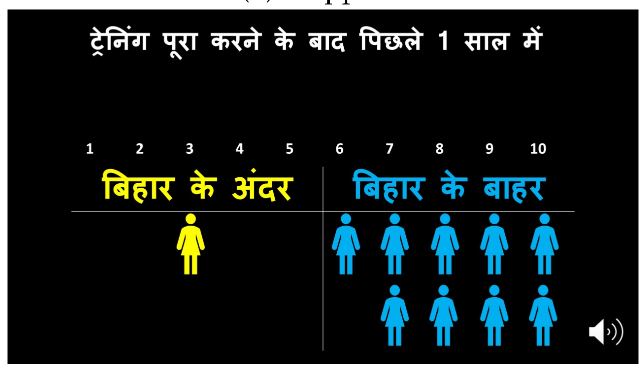
(a) Snippet 1



(b) Snippet 2

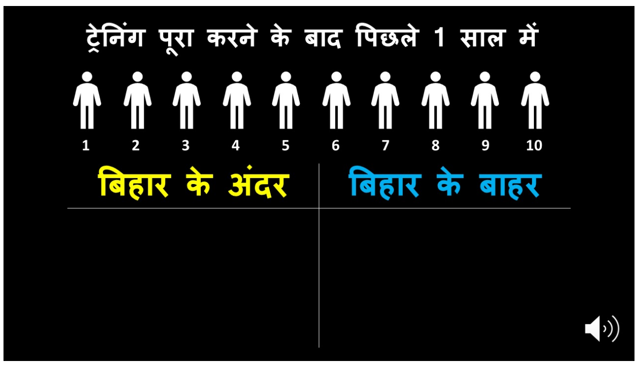


(c) Snippet 3

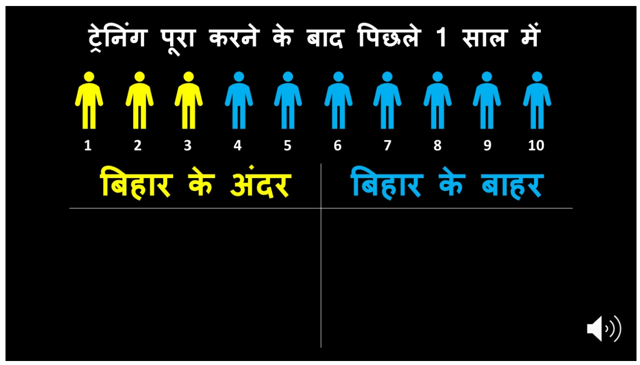


(d) Snippet 4

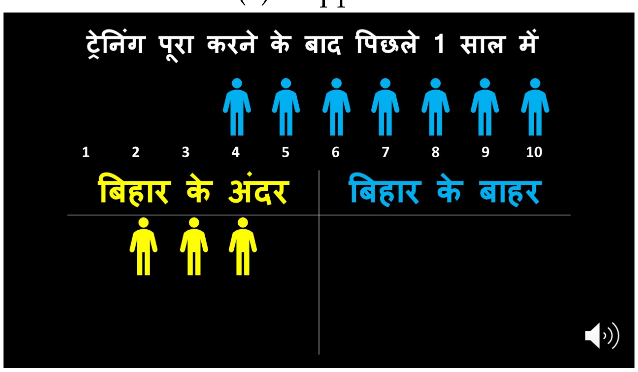
Figure A2: Location Intervention Video Snippets (Male)



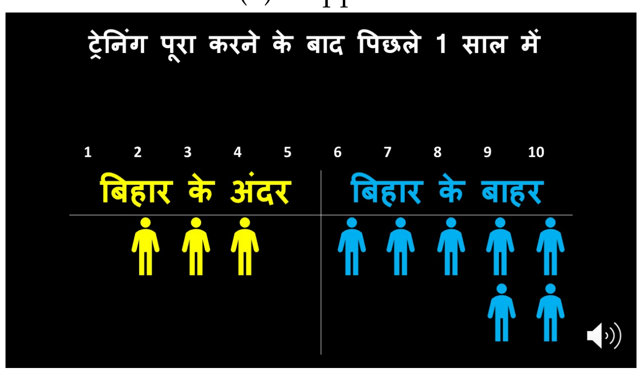
(a) Snippet 1



(b) Snippet 2

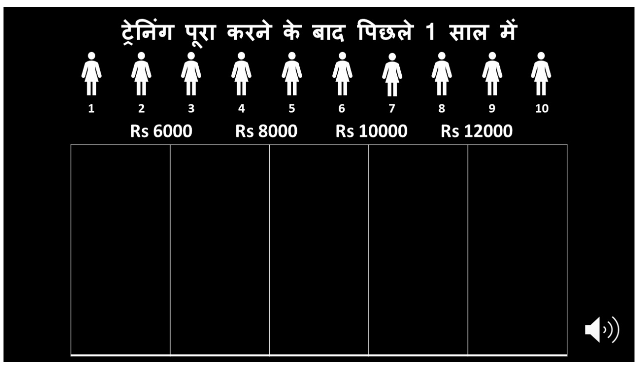


(c) Snippet 3

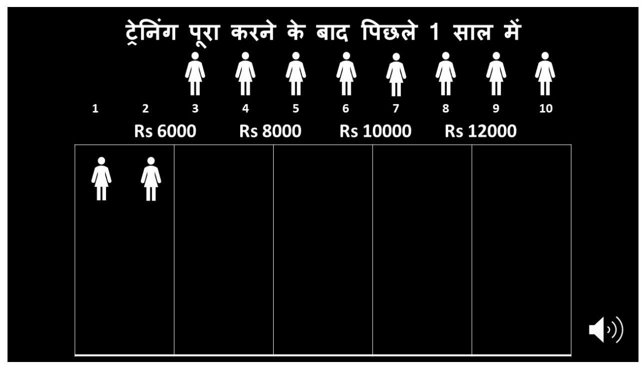


(d) Snippet 4

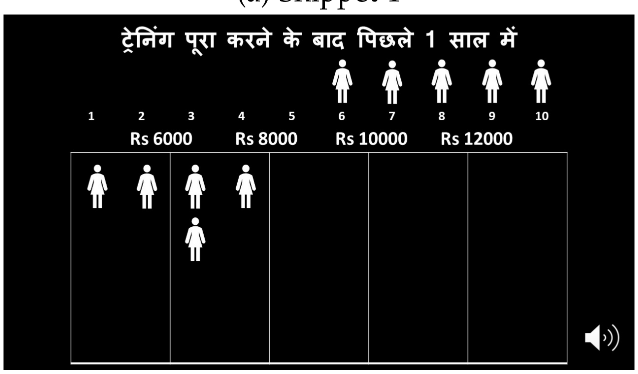
Figure A3: Salary Intervention Video Snippets (Female)



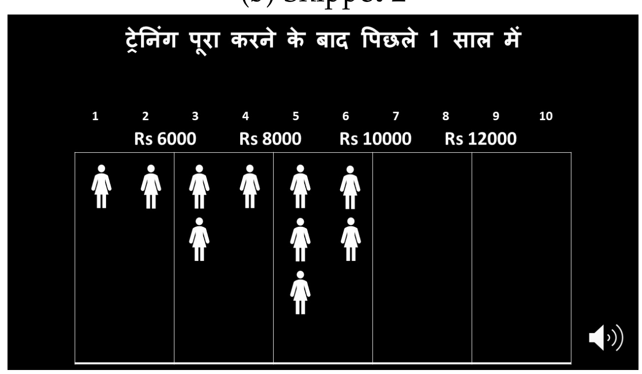
(a) Snippet 1



(b) Snippet 2

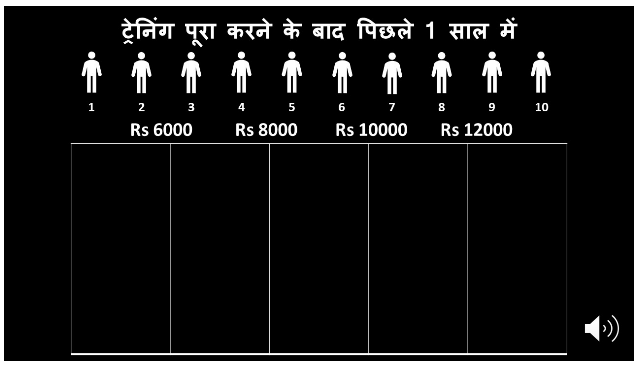


(c) Snippet 3

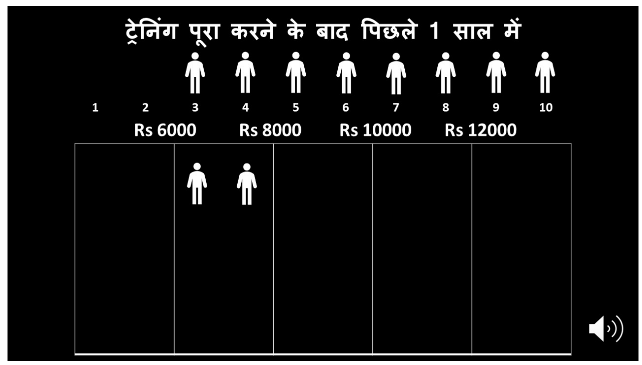


(d) Snippet 4

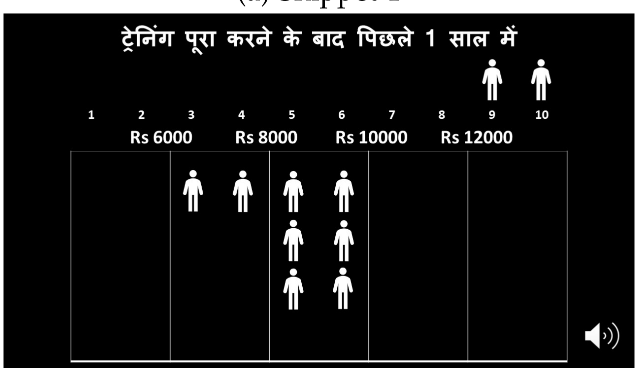
Figure A4: Salary Intervention Video Snippets (Male)



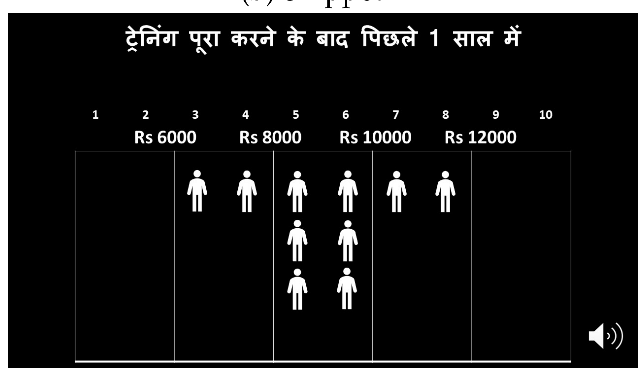
(a) Snippet 1



(b) Snippet 2

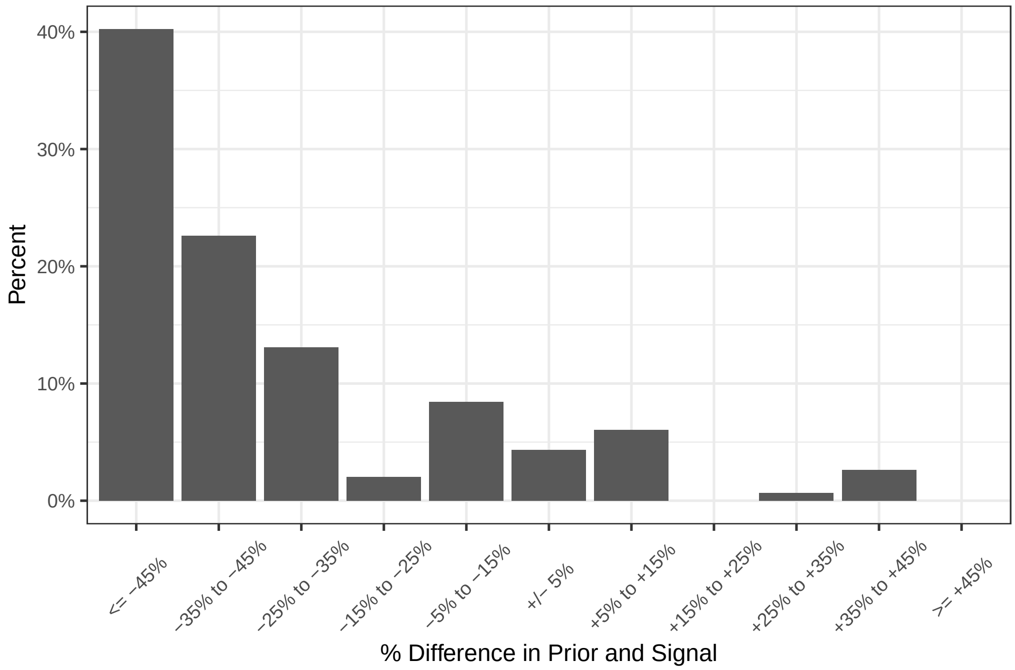


(c) Snippet 3

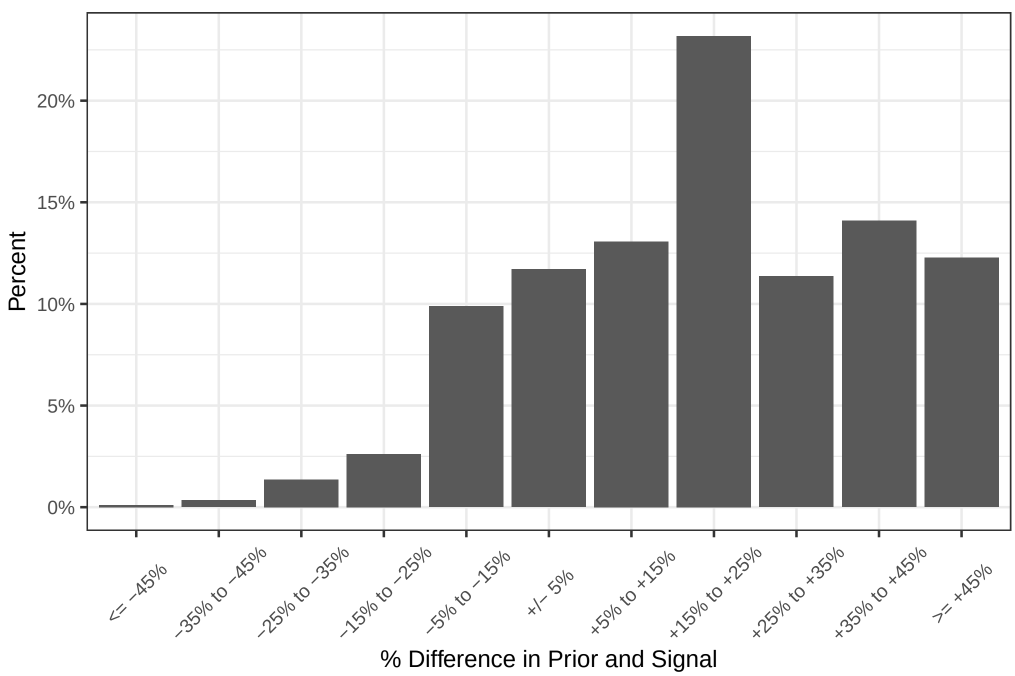


(d) Snippet 4

Figure A5: Misperceptions in Prior Beliefs

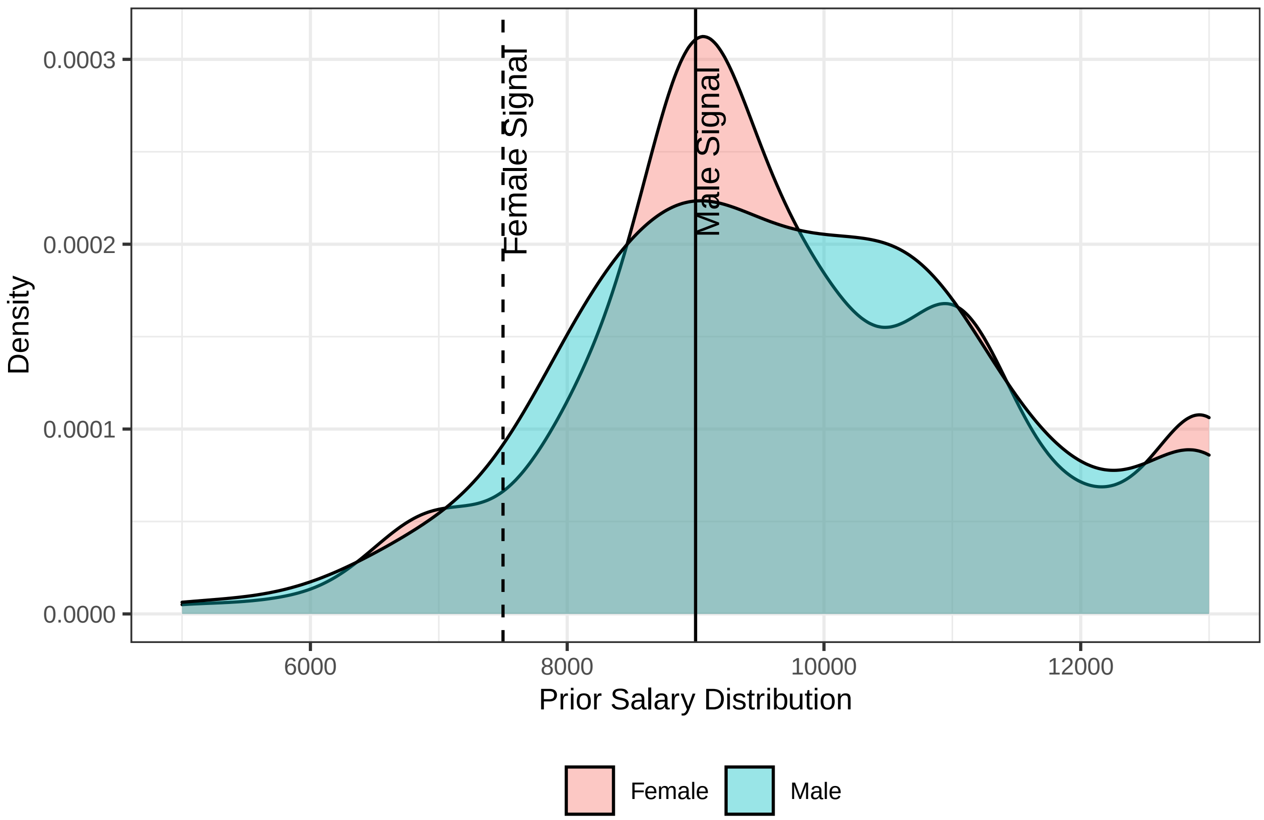


(a) Location (Candidates Outside State)



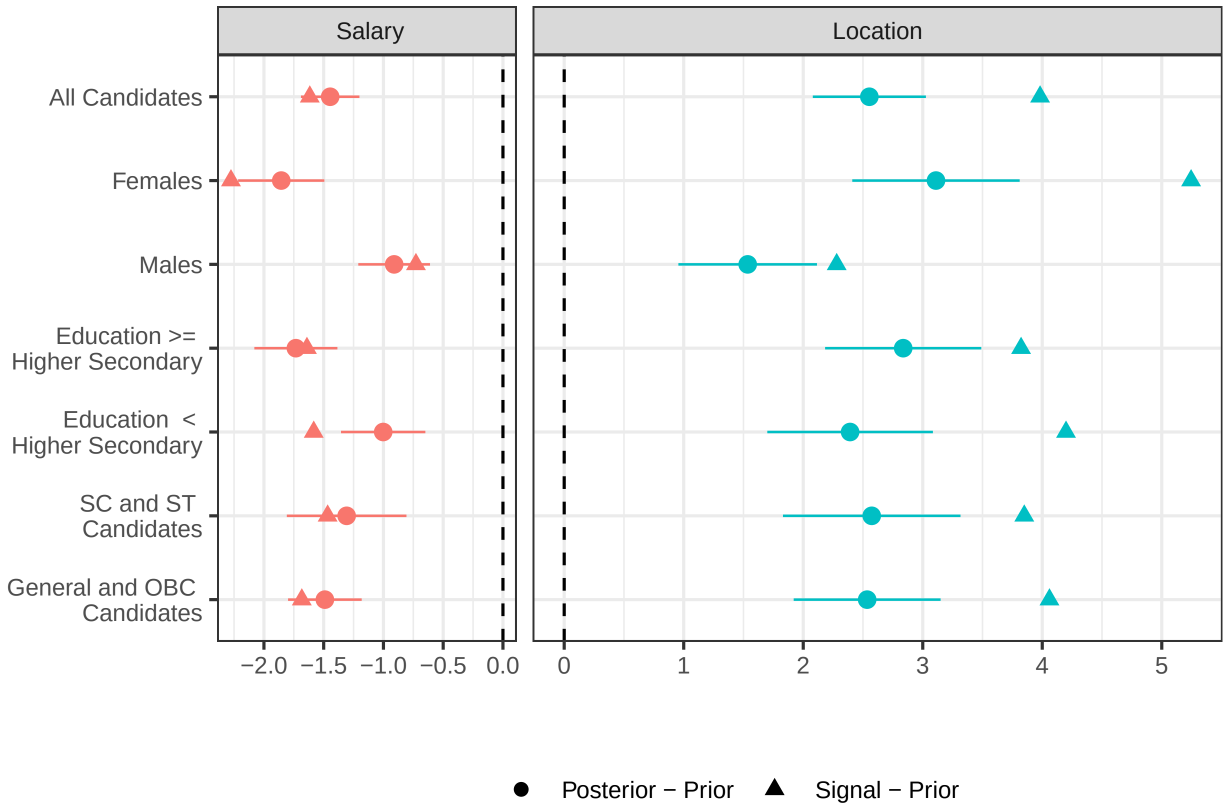
(b) Salary (Earnings Distribution Mean)

Figure A6: Prior Salary Distribution by Gender



Notes: The figure shows prior salary distribution by gender. The vertical line shows the truth/signal by gender. The actual salary for males rests at the 38th percentile of the prior salary distribution. For the females, the signal intersects the prior distribution at the 8th percentile.

Figure A7: Heterogeneity in Labor Market Beliefs



Notes: The figure shows heterogeneity in treatment effect for the salary and location intervention for sub-samples by gender, education levels, and social category of candidates. The circles and error bars show the point estimate and 95% CI on the indicator variable for the salary treatment (red color) and the location treatment (green color) regressed on the outcome variable: posterior - prior. The triangle shows the average gap between the signal and the prior. Posterior/Prior for salary is the earnings distribution mean calculated using the number of candidates in each bin. Posterior/Prior for location is the number of candidates outside state. The negative x-axis is scaled by 1000.

B Video Transcripts

B.1 Introduction Video

Voiceover: In the households of the village where there was not much enthusiasm so far, today there is hope. The young men in the rural areas, and especially the young women of the villages, who had never imagined their future outside the threshold of their houses, are today dreaming big and giving wings to their dreams because of their skills and self-confidence.

Now they are getting jobs in the organized work sector of big and metro cities.

Now happiness and smile never leaves their faces.

For lakhs of 15-35 years old rural youth, Indian Government has initiated this Rural Skill Development program.

To bring the youth from rural areas to the best training institutes and companies, this program is run on a public private partnership model.

Youth from rural areas of this country are brought and given free of cost training. Arrangements are also made for their free of cost boarding and lodging.

During the training, candidates are given books and uniform as well in the DDUGKY program.

DDUGKY program has opened lakhs of such opportunities for young men and women across this country, so that it has enabled them to write their own future with their own hands."

Female candidate: "I come from a poor family. Our family works on the farms and I have studied while working on the farms myself. My parents have taught me with great difficulty. I got to know about this free of cost training, DDU-GKY. I enquired where to get the form for this training and where is this happening. They called me that we have to leave for ranchi. . . . The facilities are good here. We had to live in hostel, the food was good.. three months we got training there. It was good, we used to have fun and play, everything was there. It feels good when we get our salaries. If we are independent people will give us importance and talk with respect.."

Male candidate: "I have my mother and father at home. We are 7 siblings, 3 brothers and 4 sisters. Before this I used to work as a daily laborer. I did not study much. I have passed my matriculation, that too with much difficulty, while working. I worked as a labor worker in a construction site where they make buildings. I worked as a helper for the masons. About DDU-GKY, they told this was a good course and they will teach us computers.."

Voiceover: "Their progressing steps towards their own brighter present are also making a stronger and developed future for India. This will turn this nation into a place of skilled individuals."

"My skill is my identity."

B.2 Intervention Video: Salary (Male)

In this video we will tell you about the monthly salary distribution of the male candidates after their training completion, in the last one year under the DDU-GKY skill development program.

Through our survey we have come to know that, after completing the training in the last one year, if 10 candidates like you got jobs, then nobody got a job for a monthly salary below Rs 6000. After completing the training in the last one year, if 10 candidates like you got jobs, then 2 male candidates got a job for monthly salary ranging between Rs 6000 to Rs 8000. Through our survey we have come to know that, after completing the training in the last one year, if 10 candidates like you got jobs, then 6 male candidates got a job for monthly salary ranging between Rs 8000 to Rs 10000. Through our survey we have come to know that, after completing the training in the last one year, if 10 candidates like you got jobs, then 2 male candidates got a job for monthly salary ranging between Rs 10000 to Rs 12000. After completing the training in the last one year, if 10 candidates like you got jobs, then nobody got a job for a monthly salary above Rs 12000.

Through this video we learn that after completing the training in the last one year, if 10 candidates like you got jobs, then nobody got a job for a monthly salary below Rs 6000, 2 male candidates got a job for monthly salary ranging between Rs 6000 to Rs 8000, 6 male candidates got a job for monthly salary ranging between Rs 8000 to Rs 10000, 2 male candidates got a job for monthly salary ranging between Rs 10000 to Rs 12000 and nobody got a job for a monthly salary above Rs 12000.

Thank you for paying attention to this.

B.3 Intervention Video: Salary (Female)

In this video we will tell you about the monthly salary distribution of the female candidates after their training completion, in the last one year under the DDU-GKY skill development program.

Through our survey we have come to know that, after completing the training in the last one year, if 10 female candidates like you got jobs, then 2 female candidates got a job for a monthly salary below Rs 6000. Through our survey we have come to know that, after completing the training in the last one year, if 10 female candidates like you got jobs, then 3 female candidates got a job for monthly salary ranging between Rs 6000 to Rs 8000.

Through our survey we have come to know that, after completing the training in the last one year, if 10 female candidates like you got jobs, then 5 female candidates got a job for monthly salary ranging between Rs 8000 to Rs 10000. Through our survey we have come to know that, after completing the training in the last one year, if 10 female candidates like you got jobs, then nobody got a job for monthly salary ranging between Rs 10000 to Rs 12000. After completing the training in the last one year, if 10 female candidates like you got jobs, then nobody got a job for a monthly salary above Rs 12000.

Through this video we learn that after completing this training in the last one year, if 10 female candidates like you got jobs, then 2 female candidates got a job for a monthly salary below Rs 6000, 3 female candidates got a job for monthly salary ranging between Rs 6000 to Rs 8000, 5 female candidates got a job for monthly salary ranging between Rs 8000 to Rs 10000, nobody got a job for monthly salary ranging between Rs 10000 to Rs 12000 and nobody got a job for a monthly salary above Rs 12000.

Thank you for paying attention to this.

B.4 Intervention Video: Location (Male)

In this video we will tell you about the job location of the male candidates after their training completion, in the last one year under the DDU-GKY skill development program.

Male candidates who got placed inside Bihar are in yellow color and male candidates who were placed outside Bihar are in blue color.

Through our survey we have come to know that, after completing the training in the last one year, if 10 male candidates like you got jobs, out of them 3 male candidates got a job inside Bihar and through our survey we have come to know that, after completing the training in the last one year, if 10 male candidates like you got jobs, out of them 7 male candidates got a job outside Bihar.

Through this video we learn that after completing the training in the last one year, if 10 male candidates like you got jobs, out of them 3 male candidates got a job inside Bihar and 7 male candidates got a job outside Bihar.

Thank you for paying attention to this.

B.5 Intervention Video: Location (Female)

In this video we will tell you about the job location of the female candidates after their training completion, in the last one year under the DDU-GKY skill development program.

Female candidates who got placed inside Bihar are in yellow color and female candidates who were placed outside Bihar are in blue color.

Through our survey we have come to know that, after completing the training in the last one year, if 10 female candidates like you got jobs, out of them 1 female candidate got a job inside Bihar and through our survey we have come to know that, after completing the training in the last one year, if 10 female candidates like you got jobs, out of them 9 female candidates got a job outside Bihar.

Through this video we learn that after completing the training in the last one year, if 10 female candidates like you got jobs, out of them 1 female candidate got a job inside Bihar and 9 female candidates got a job outside Bihar.

Thank you for paying attention to this.

C Consent Forms

Baseline Survey Hello. I am conducting a survey on behalf of Warwick University. We are carrying a research study to improve the training programme DDU-GKY. This survey will take less than 15 minutes. This interview is voluntary: you are free to decide not to participate in this survey and you have the right not to answer any question. Your personal information will be stored securely and your answers will be anonymized at the end of the study. Once anonymized your data may be kept for future research. Please read the instructions carefully and ask any question you may have. If you have questions later you can call ABC at this number XXXXXXXXX.

Followup Survey Hello. I am conducting a survey on behalf of Warwick University. We are carrying a research study to improve the training programme DDU-GKY for which you had participated in a survey work earlier at mobilisation camp. You had provided your phone number in that survey. We would like to ask a few questions and it will not take more than 5 minutes. You will be provided with a mobile recharge between Rs 30 and Rs 50 depending on your telecom provider if you choose to participate in the survey within 10 days. This interview is voluntary: you are free to decide not to participate in this survey and you have the right not to answer any question. Your personal information will be stored securely and your answers will be anonymized at the end of the study. Once anonymised your data may be kept for future research. If you have any questions, please feel free to ask. If you have questions later you can call me at this number.

D Pre-Analysis Plan

D.1 Introduction

India, like other developing countries, suffers from low productivity of labour (see the IGC Evidence paper by Bloom et al., 2014). Training the labour force is the primary policy for increasing skills and labour productivity. However, the literature has shown that designing successful training programmes is difficult (Blattman and Ralston, 2015; McKenzie, 2017). In many instances, they suffer from low take-up and high attrition rates, which plague the impact on final outcomes. In this research, we aim to study the what role information about job prospects (wages and location) play in enrollment and completion, in the case of a large training programme in India.

"Deen Dayal Upadhyaya Grameen Kaushal Yojana" (DDU-GKY) is one of India's major and most prominent skills and job creation schemes, launched in 2014. The scheme is residential, attracting candidates from remote villages, and mandates that each trainee is offered a job at the end of the training. The scheme follows a Public-Private Partnership model, where registered private sector project implementation partners (PIAs) bid for government funds, and plan and implement skills training and job placement programs, targeting rural youth from poor families (DDU-GKY Programme Guidelines, 2016).

DDU-GKY is implemented in 29 states and union territories. It currently has over 1426 projects being implemented by over 649 partners, in more than 552 trades from 52 industry sectors. Over 920,000 candidates have been trained and over 490,000 candidates have been placed in jobs as of December 30, 2019.[19]

Based on our analysis of administrative data and qualitative interviews with PIAs in Bihar and Odisha, we observe that dropouts during training, as well as from the post-training placement jobs, are of serious concerns. We hypothesize that candidates are misinformed about the objectives of the programme they enroll in, and about the jobs they will be offered post-training. This mismatch leads to drop out when they learn more about programme and jobs.

We conduct an experimental study using a randomised controlled trial (RCT), to evaluate how information about two crucial aspects of the jobs (distribution of wages and job location) affect training enrollment and completion decisions. Candidates are enrollled in training programmes through mobilisation camps organized at the village/panchayat/block level. The control group receives basic information about the training programme, accom-

[19] `http://ddugky.gov.in/content/about-us-0` accessed on February 25, 2020.

modation and food facilities during the training. The three treatment groups receive information on distribution of wages (1), distribution of job location (2) and both distribution on wages and job location (3). These interventions attempt to reduce the mismatch between candidates' expectations and the employment opportunities offered by the programme.

We conduct two follow-up surveys, one week and four week after the baseline survey to measure the candidate's expectation about wages and job location and intention to enroll in training program. We then follow these candidates using the administrative data to see if they have actually enrolled in any training batch as well as training completion and employment outcomes. We expect the treated trainees to update their expectations about salary and job location. The information intervention carried out at the time of mobilisation will filter out candidates whose job expectations do not match with what can be achieved from the training programme. The remaining treated trainees are expected to experience lower attrition conditional on enrollling.

If the interventions proved to be successful, they could easily be implemented by local training centres in a first step, and by training centres in other states in a second step. If these interventions were unsuccessful, we would collectively learn that the (mis)information about employment opportunities is not key to the mismatch problem in this context and should look for other mechanisms to explain dropout from training and jobs.

The setting for this research is in the state of Bihar, one of India's poorest states, where caste- and sex-based hierarchy is pronounced. DDU-GKY is explicitly targeted towards females, scheduled castes and scheduled tribes (SC-STs), and we are particularly interested in the impact of our interventions on these marginalised groups, as they might suffer from larger informational deficits about labour markets.

D.2 Experimental design

D.2.1 Background

From qualitative interviews, we have identified potential trainees are mis-informed about two important aspects of job opportunities: (i) the wages offered; (ii) location of job (inside or outside the state). These incorrect expectations could step from:

- The mobiliser of the PIAs might be communicating incorrect information to enroll candidates and achieve personal targets.
- The JRP (job resource person) in-charge of organizing mobilisation camps for the PIAs might be mis-informing candidates about training prospects.

- Other community members, past candidates etc. might also be a source of false information.

We hypothesise that the lack of truthful information about placement jobs, leads to mismatch between the trainee's expectations, and what the programme offers, at the enrollment stage. An information intervention (trying to tilt agents' expectation towards their true values) at the stage of mobilisation:

- will have mixed results on the enrollment rate, depending on whether individual trainees are over- or under-optimistic on each aspect of the job.
- should unambiguously increase the placement rate, conditional on enrolling.

D.2.2 Research Questions

We aim to address the following questions:

1. Do individuals update beliefs in response to the information interventions?
2. What share of enrollment and non-enrollment can be thought to be the result of well-informed decisions? In this case, which factors explain the most heterogeneity in the population (between those who enroll and those who don't)?
3. How much do candidates trade off location and salary, i.e. how much do they value proximity of the job offered to home? How does this value differ across socio-economic groups.

D.2.3 Information Intervention

The JRP (job resource person) organizes mobilisation camps at the village/panchayat/block level in collaboration with the PIAs by inviting candidates who might be interested in the skill training programme. The PIAs send their mobilisers to each camps to provide information about the training centre, trade and batch start date. In these camps, each candidate is invited to take part in the survey and is assigned to one of the four intervention arms.

- Control: Candidate gets to see a basic informational video about DDU-GKY program. The video provides a glimpse of the training centre, accommodation and food facilities, and classrooms. Two placed candidates happily describe improvement in their lives from the training. At no point the video provides any information on job location or the wages offered.

- Treatment Location: Candidate watches the basic information video and an additional video which provides information on the distribution of job location for past DDU-GKY candidates. The distribution of job location is elicited the share of candidates out of 10 who get a job inside Bihar and outside Bihar.
- Treatment Salary: Candidate watches the basic information video and an additional video which provides information on the distribution of wages for past DDU-GKY candidates. The distribution of wages is elicited the share of candidates out of 10 who get a monthly salary in the following bins: less than Rs 6000 per month, Rs 6000 - 8000 per month, Rs 8000 - 10000 per month, Rs 10000 - Rs 12000 per month and more than Rs 12000 per month.
- Treatment Salary + Location: Candidate watches the basic information video and two additional videos which shows the distribution of wages and job location for past DDU-GKY trainees.

The distribution of wages and location of past DDU-GKY trainees has been obtained from the surveys of a parallel project in the same state. Since the wages and job location differ across male and female candidates, the true distribution is tailored to the gender of the candidate. We expect that: (i) candidates who are over-optimistic about the placement prospects to not enroll in the program, (ii) candidates who were under-optimistic to begin with will enroll in the program and (iii) conditional on enrolling with correct expectations, the training completion should unambiguously increase. The randomization is done at the individual level, we expect a sample size of 850 candidates.

D.2.4 Power calculations

We consider a power of 80% and a significance level of 5%. We ran simulations to obtain the effect size to detect significant difference between the treatment and control on the outcome variable (admission - enrollment at the training centre). A point estimate of 11 pp on the updating in salary distribution and -7 pp on the updating in location distribution will be enough to attain a power of 80%. The power at the second stage still holds with at most 4 pp change in the β_s & β_l at the first stage.

The table below presents changes in power calculations as we vary the effect size (β_s and β_l) on the treated candidates for the equations presented in section 4.2.2 (Columns 3 & 4) and section 4.2.3 (Column 5 & 6). The coefficients used in the simulation exercise comes from the pilot data (denoted by $*$) or from the baseline data (denoted by #) and are as follows:

- Number of observations: 701#
- Salary:
  - prior mean: 9,500#
  - prior standard deviation: 1,700#
  - posterior standard deviation on error: 1,700*
  - true signal: 8,400#
  - Regression parameters (Section 4.2.1): $\beta_1^s = -900$, $\beta_2^s = 0.2$, $\beta_3^s = 0.3^*$
- Location:
  - prior mean: 4#
  - prior standard deviation: 2.7#
  - posterior standard deviation on error: 3*
  - true signal: 8#
  - Regression parameters (Section 4.2.1): $\beta_1^l = 0$, $\beta_2^l = 0.6$, $\beta_3^l = 0.3^*$
- Probability of Enrollment:
  - prior standard deviation: 0.28#
  - posterior standard deviation: 0.25*
- Admission:
  - posterior standard deviation: 0.34*

The estimation is done using ordinary least squares regression. There might be additional noise between the posteriors at the baseline and each followup which has not been considered. The first stage (updating of beliefs) generally has a high power. The results for second stage are presented for probability of enrollment (Columns 3 & 4) and admission (Columns 5 & 6) by varying the β_s in first four rows and varying the β_l in the last four rows. The effect size to detect significant difference between the treatment and control with at least 80% power are obtained if $\beta_s = 0.11$ and $\beta_l = -0.07$. The power at the second stage still holds with at most 4 pp change in the β_s & β_l at the first stage.

Table A13: Power calculation simulations across effect size

	First Stage		Second Stage			
			P(Enrollment)		Admission	
	Salary	Location	Salary	Location	Salary	Location
	(1)	(2)	(3)	(4)	(5)	(6)
$\beta_s = 0.08$ & $\beta_l = -0.09$	97.1%	94.6%	84.1%	99.6%	60.8%	95.3%
$\beta_s = 0.10$ & $\beta_l = -0.09$	97.1%	94.6%	95.6%	99.6%	77.7%	95.3%
$\beta_s = 0.11$ & $\beta_l = -0.09$	97.1%	94.6%	98.1%	99.6%	85.1%	95.3%
$\beta_s = 0.12$ & $\beta_l = -0.09$	97.1%	94.6%	99.2%	99.6%	91.0%	95.3%
$\beta_s = 0.10$ & $\beta_l = -0.05$	97.1%	94.6%	95.6%	80.0%	77.7%	55.5%
$\beta_s = 0.10$ & $\beta_l = -0.06$	97.1%	94.6%	95.6%	91.0%	77.7%	71.9%
$\beta_s = 0.10$ & $\beta_l = -0.07$	97.1%	94.6%	95.6%	96.6%	77.7%	84.1%
$\beta_s = 0.10$ & $\beta_l = -0.09$	97.1%	94.6%	95.6%	99.6%	77.7%	95.3%

D.3 Data collection

D.3.1 Collection process

Our research relies on:

- Primary data collected from three rounds of surveys of potential trainees from mobilisation camps conducted across multiple districts in Bihar.
- Data from the management information system (MIS) data from Bihar Rural Livelihood Promotion Society (BRLPS).

Surveys. All surveys are administered on tablets using questionnaires designed on Survey CTO platform. The baseline survey is administered in face-to-face interviews. The two followup surveys are administered using phone interviews. The camp activity survey for each mobilisation camp is done by the enumerator.

- Camp activity survey: This survey is conducted for each mobilisation camp by the data collector. The objective is to collect information on the topic covered during the information conveyed by the PIA mobiliser and the JRP.
- Baseline survey: This survey is being administered to all participants in the mobilisation camps, after the trainees have received information from the JRP and PIA mobiliser. Data collectors administer the baseline questionnaire in a face-to-face interview sessions with individual trainees. The baseline questionnaire is custom designed

to collect information and disseminate the intervention according to the treatment assignment and gender of the candidate. The questionnaire captures probability of joining the training, prior and posterior distribution of wages and job location, expected and counterfactual earnings 1 year post training and socio-economic characteristics of the candidate.

- Followup surveys: This telephonic-survey interview is conducted with the trainees after one week and four weeks of the baseline survey. The objective is to collect information about the posterior distribution of wages and job location, expected and counterfactual earnings and if the candidates have made a decision to join the program. The details are provided below:

Administrative data. We will be able to match the survey data with the BRLPS administrative data on DDU-GKY. Crucially, these data contain the dates when each trainee enrolled, dropped out (if they did), graduated, and enrolled in the placement job. The administrative data relies on reporting from training providers to the state administration, which are sometimes incomplete and could potentially be erroneous.

D.3.2 Outcomes

Main outcomes. The main outcomes are measured using baseline and two followup survey data.

1. Posterior distribution of wage and job location at the end of training.
2. Probability of enrollment at the training centre.
3. Admission at the training centre.

As a robustness check, we will report results on enrolment, training completion and placement based on administrative data.

Secondary outcomes. Secondary outcomes are measured in the baseline and the two followup surveys. They will be used for interpretation:

- Expected earnings, occupation, location in 12 months from baseline if you enroll.
- Expected earnings, occupation, location in 12 months from baseline if you don't enroll.

- Visit to the training centre.
- Posterior probability to join the training centre.

D.3.3 Control variables

We measure a range a variables at baseline in order to improve the precision of our estimators.

- Sex.
- Age.
- Social category.
- Religion.
- Highest education level completed.
- Prior probability to join the training centre.
- Prior distribution of wages and job location.
- BPL card / RSBY card / SHG Membership / MGNREGA participation in household.
- District, block and panchayat of candidate.
- PIA name.
- PIA district and block.

D.3.4 Other variables

We measure a range a variables through a common survey of the mobilisation camp.

- Mobilisation camp district and block.
- Camp level (village/panchayat/cluster/block)
- PIA mobiliser presence.
- Topics discussed by the PIA mobiliser
- Topics discussed by the JRP
- Number of candidates who heard the speech of JRP and mobiliser.

- Target gender of the camp.
- When were the candidates informed about the camp.
- Presence of past DDU-GKY dropout candidate.

D.4 Econometric analysis

We estimate models using observations on individuals present at the baseline survey. An individual i is assigned to either Treatment Salary T_i^s or a Treatment Location T_i^l or both and has a vector of characteristics X_i (control variables).

D.4.1 Balance

To check that our randomisation achieved balance between treatment and control at baseline, we will estimate for each control variable X_i':

$$X_i = \beta_s T_i^s + \beta_l T_i^l + \beta_{sl} T_i^s \times T_i^l + \varepsilon_i$$

We will then test the null of no difference between the treatment groups and control group ($\beta_s = 0$, $\beta_l = 0$ and $\beta_{sl} = 0$). We will correct for multiple hypothesis testing, controlling for false discovery rates.

D.4.2 Beliefs

$$Posterior_{ic}^j - Prior_{ic}^j = \gamma^j T_{ic}^j + X_{ic}'\alpha + \delta_c + \varepsilon_{ic}$$

$$Posterior_{ic}^j - Prior_{ic}^j = \beta_1^j T_{ic}^j + \beta_2^j (Signal^j - Prior_{ic}^j) + \beta_3^j (Signal^j - Prior_{ic}^j) \times T_{ic}^j + X_{ic}'\alpha + \delta_c + \varepsilon_{ic}$$

$$j \in \{s, l\}$$

$Prior_{ic}^j$ and $Posterior_{ic}^j$ measures the mean of candidate i's prior and posterior distributions for salary (j = s) and location of job (j = l) respectively at the end of training attending the mobilisation camp c. $Signal^j$ is the true information signal for salary (j = s) and location of job (j = l). γ^j is the intention-to-treat estimates, the quantity of interest in our setting. β^s and β^l capture the heterogenous effect by priors of the candidates. We will use post-double-selection lasso for variable selection. We will correct for multiple hypothesis testing, controlling for false discovery rates. The estimation will also be done by gender since the $Signal^j$ varies by gender.

It maybe the case that the information interventions have cross-interaction effects. For example, the information intervention on distribution of job location might have an effect on the posterior distribution of wages. If it is indeed the case, then the above analysis will be done with both treatment arms in the right-hand side of the equation.

D.4.3 Probability of Enrollment

First Stage:

$$Posterior^j_{ic} - Prior^j_{ic} = \beta^j_1 T^j_{ic} + \beta^j_2(Signal^j - Prior^j_{ic}) + \beta^j_3(Signal^j - Prior^j_{ic}) \times T^j_{ic} + X'_{ic}\alpha + \delta_c + \varepsilon_{ic}$$

Second Stage:

$$P(Enrollment)^{Posterior}_{ic} - P(Enrollment)^{Prior}_{ic} = \beta_l(Posterior^l_{ic} - Prior^l_{ic})$$

$$+\beta_s(Posterior^s_{ic} - Prior^s_{ic}) + X'_{ic}\alpha + \delta_c + \varepsilon_{ic}$$

$$j \in \{s, l\}$$

T^s_i and T^l_i are indicator variables if the candidate i received the intervention on salary and location respectively. Prior distributions are measured during the baseline before the interventions. Posterior distributions are measured during the baseline after the intervention and two followup surveys. The outcome variable $P(Enrollment)$ measures probability to enroll at the training centre.

D.4.4 Admission

First Stage:

$$Posterior^j_{ic} - Prior^j_{ic} = \beta^j_1 T^j_{ic} + \beta^j_2(Signal^j - Prior^j_{ic}) + \beta^j_3(Signal^j - Prior^j_{ic}) \times T^j_{ic} + X'_{ic}\alpha + \delta_c + \varepsilon_{ic}$$

Second Stage:

$$y_{ic} = \beta_l(Posterior^l_{ic} - Prior^l_{ic}) + \beta_s(Posterior^s_{ic} - Prior^s_{ic}) + X'_{ic}\alpha + \delta_c + \varepsilon_{ic}$$

$$j \in \{s, l\}$$

T_i^s and T_i^l are indicator variables if the candidate i received the intervention on salary and location respectively. Prior distributions are measured during the baseline before the interventions. Posterior distributions are measured during the baseline after the intervention and two followup surveys. The outcome variable y_{ic} is a dummy which measures if the candidate reported taking admission at the training centre. The estimation will be done using both OLS and logit regressions.

D.4.5 Attrition

Survey data may suffer from attrition, and administrative data may have missing information. We will check that the attrition rate (missing data) is not different between treatment and control, and test that our results are robust to using Lee (2009) bounds.

D.4.6 Heterogeneity

We will consider the following dimensions of heterogeneity:

- gender.
- social background (Scheduled Caste / Scheduled Tribes vs not).
- education (completed higher secondary vs not).

There are several reasons why people of different sex, social category, and education would have different treatment effects. Women, SC-ST, and individuals with low levels of education, have lower options outside of the programme. We expect a positive effect of the information interventions on changes in expectations, enrollment and training completion.

In addition to these dimensions, we will also use causal forest method on all variables listed as other variables and control variables for estimating heterogeneous treatment effects. This method allows one to pick the characteristics that are most relevant for explaining the heterogeneity in treatment effects. The estimation will be done using both first stage and second stage regressions described above.

D.4.7 Spillover Effects

The information interventions can have spillover effects if candidates talk to each other after the delivery of interventions. We expect limited across gender interactions. We will explore these effects by comparing the respondents posterior distributions between the followup and the baseline. Specifically:

$$Posterior_{ic}^{j,F} - Posterior_{ic}^{j,B} = \beta_1^j T_{ic}^j + \beta_2^j S_c^j + \beta_3^j S_c^j \times T_{ic}^j + X_{ic}'\alpha + \delta_c + \varepsilon_{ic}$$

$Posterior_{ic}^{j,F}$ and $Posterior_{ic}^{j,B}$ measures the mean of candidate i's posterior distributions for salary (j = s) and location of job (j = l) at followups (F) and baseline (B) respectively attending the mobilisation camp c. The $Posterior_{ic}^{j,B}$ is being measured right after the intervention but still during the baseline survey and is unaffected by peer interactions. S_c^j measures the share of treated candidates in a peer group (defined as mobilisation camp x panchayat) who received a treatment $j \in \{s, l\}$.

E Deviations from Pre-Analysis Plan

While our empirical approach largely follows the pre-analysis plan, we made two key modifications to refine the estimation strategy: (i) we adjusted the first-stage specification to jointly estimate the effects of both treatment arms, and (ii) we refined the outcome variable in the second stage to better capture changes in enrollment behavior relative to prior expectations.

Modification of the First-Stage Specification In the pre-analysis plan, we outlined separate first-stage equations for each treatment dimension ($j \in s, l$), estimating the impact of treatment on belief updating independently for salary and location. In the final analysis, we modified this approach by estimating a single first-stage equation that includes both treatment arms jointly. This adjustment allows us to account for potential interactions between the two treatment dimensions and ensures that our estimates of belief updating are not confounded by omitting the influence of the other treatment. The revised first-stage equation retains the key elements from the PAP, including the direct effect of treatment, the role of prior beliefs, and the interaction between the two, but integrates them into a unified framework that improves efficiency and consistency in estimation.

Refinement of the Outcome Variable in the Second Stage The pre-analysis plan specified a second-stage equation where the dependent variable was a binary indicator for enrollment in the training program. In our final specification, we refined this outcome measure by modeling the change in enrollment behavior relative to prior expectations. Specifically, we use the difference between actual enrollment decisions (I(Enrollment)$_{i,c}^{Posterior}$) and baseline self-reported probabilities of enrollment (P(Enrollment)$_{i,c}^{Prior}$). This approach better captures how individuals update their enrollment decisions in response to new information and mitigates potential biases that arise from pre-existing differences in enrollment probabilities. By modeling the deviation from prior expectations, we provide a clearer measure of how belief updating influences decision-making, rather than focusing solely on absolute enrollment rates.

These modifications align with our original research objectives while improving the robustness of the empirical strategy. They allow us to better isolate the causal effects of belief updating on enrollment decisions and ensure that our estimates account for both treatment dimensions simultaneously.

Additional Analysis on Spillover Effects Table A12 was added at the request of a referee and was not part of the original pre-analysis plan. This additional exploratory analysis investigates potential spillover effects on beliefs—both priors and posteriors—during the baseline and follow-up surveys. While not pre-specified, this analysis provides further reassurance that the main treatment effects are not driven by cross-group information diffusion.