I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14427

Can Information about Jobs Improve the Effectiveness of Vocational Training? Experimental Evidence from India

Bhaskar Chakravorty
Wiji Arulampalam
Apurav Yash Bhatiya
Clément Imbert
Roland Rathelot

MAY 2021 (THIS VERSION: APRIL 2023)

DISCUSSION PAPER SERIES

IZA DP No. 14427

Can Information about Jobs Improve the Effectiveness of Vocational Training? Experimental Evidence from India

Bhaskar Chakravorty
University of Warwick

Wiji Arulampalam
University of Warwick and IZA

Apurav Yash Bhatiya
University of Birmingham

Clément Imbert
University of Warwick, BREAD, CEPR, EUDN and J-PAL

Roland Rathelot
University of Warwick CEPR and J-PAL

MAY 2021 (THIS VERSION: APRIL 2023)

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Can Information about Jobs Improve the Effectiveness of Vocational Training? Experimental Evidence from India*

We use a randomized experiment to evaluate the impact of providing richer information about prospective jobs to vocational trainees on their employment outcomes. The setting of the study is the vocational training program DDU-GKY in India. We find that including in the training two information sessions about placement opportunities make trainees 18% more likely to stay in the jobs in which they are placed. We provide suggestive evidence that the effect is driven by improved selection into training: as a result of the intervention, trainees that are over-optimistic about placement jobs are more likely to drop out before placement.

JEL Classification: J24, J61, M53

Keywords: vocational training, on-the-job-training, dropout, rural development, skills, job placement

Corresponding author:
Wiji Arulampalam
Department of Economics
University of Warwick
Coventry, CV4 7AL
United Kingdom
E-mail: wiji.arulampalam@warwick.ac.uk

* We thank Gaurav Chiplunkar, Santhosh Mathew, Kate Orkin, Rohini Pande, Simon Quinn, Chris Woodruff, as well as participants in seminars at ISI, PSE, Oxford CSAE, Goettingen, NEUDC, and 2021 BREAD Conference on Behavioral Economics and Development for useful comments at several stages of the project. We thank Maximilian Lohnert for amazing research management, Rishi Kishore, Avash Kumar, Kritarth Jha, Poonam Panda, Todor Tochev for excellent research assistance. This project would not have been possible without the invaluable collaboration of BRLPS, JSLPS and the Ministry of Rural Development. Funding from the following organizations is gratefully acknowledged: International Growth Centre, EQUIP-ESRC, CAGE and Economics Department at the University of Warwick. The revised version of the paper includes Apurav Yash Bhatiya as an additional co-author. This trial was registered with the American Economic Association registry for randomized control trials, under the title "The right person for the right job: Can information about job prospects reduce dropout in a large Indian training program?", and the trial number AEARCTR-0003611. The ethical approval was obtained from the Research Ethics Committee of the University of Warwick (HSSREC 56/17-18), and IFMR (IRB00007107).

1 Introduction

Youth unemployment and underemployment are major issues for developing countries. Although vocational training programs have been found to be generally effective in promoting employment in some contexts (Alfonsi et al., 2020; Maitra and Mani, 2017), evidence is mixed (Betcherman, 2004; Blattman and Ralston, 2017; McKenzie, 2017). In many instances, high attrition from programs limit their impact on employment outcomes. For example, Heckman et al. (2000) report dropout as high as 79% in U.S. training programs. Developing-country examples include Hirshleifer et al. (2016) for Turkey, Card et al. (2011) for Dominican republic, and Cho et al. (2013) for Malawi. One potential reason for this attrition is the mismatch between youth expectations and the jobs available to them. In training programs that also do placement, this mismatch is potentially easy to address by better information about placement opportunities.

We examine this question in the context of the vocational training program DDU-GKY (Deen Dayal Upadhyay Grameen Kaushalya Yojana), one of the largest vocational training programs in the world, launched in 2014 by the Indian government for the rural youth. The scheme is implemented in a public–private partnership mode, whereby registered private sector partners or project implementation agencies (PIA) plan and implement skills training and place program participants. Although the program guarantees placement to every trainee, only about 60% of the 1.3 million DDU-GKY trainees have been placed so far.[1] We build on three years of collaboration with the agencies in charge of DDU-GKY in two of the poorest states of India (Bihar and Jharkhand). We developed two short information sessions that provided details of the placement opportunities (e.g. job title, company name, location, compensation) which were known to the training provider but not the trainees. We report on a randomized experiment that evaluates the effect of these sessions on training completion and job placement.

A simple conceptual framework suggests that providing information about prospective jobs can improve placement outcomes of vocational trainees in two ways. First, some trainees may be over-optimistic about placement prospects, and participate in the training even though their outside options are better than the placement jobs provided by the scheme, while others maybe over-pessimistic, and drop out too early. Better information will lead the over-optimistic trainees to leave the training and the over-pessimistic to stay. This *selection* channel implies that the intervention would increase the probability of staying in the job conditional on placement but have an ambiguous effect on average training completion and placement. Second, knowing about the details of the placement jobs may make it eas-

[1] Official statistics from http://ddugky.gov.in/ accessed on 31st March 2023.

ier for trainees to transition into employment. This *job readiness* channel implies that the intervention increases the value of the job, decreases dropout, and increases placement.

We find that trainees in the treatment group were on average 18% more likely to stay in the jobs in which they were placed, but we do not find a significant effect on dropout or placement on average. These findings tend to support the selection rather than the job-readiness channel. A direct way to test for the mechanisms is to estimate the effect of information on trainees' expectations about the placement jobs. We leverage post-intervention data on trainees expectations, and find some evidence that trainees in the treatment group are on average less optimistic about their likelihood of taking-up and staying in placement jobs. However, these results are not fully conclusive because the data were collected a few months after the first information session, and we were unable to measure expectations for trainees who had updated their expectations negatively and dropped out.

To provide more evidence on the mechanisms, we test for heterogeneous effects of our treatment along three dimensions: gender, education, and caste. In terms of gender, we find the treatment effects to be entirely concentrated on men, for whom the treatment increased the probability of staying for at least five months in the placement job by 55%. In contrast, there was no effect on women, who in the control group were much more likely to complete the training and take up placement jobs than men. This is consistent with a selection channel, because women in this context have far worse labor market opportunities than men outside of the program, so mismatched expectations are likely to be a bigger issue for men. Turning to heterogeneity by education levels, we find the intervention to have opposite effects on less educated trainees relative to more educated ones. For the less educated trainees, who in the absence of the intervention had a higher dropout relative to more educated ones, the intervention reduced dropout by 35%. For more educated trainees, however, the treatment increased dropout by 50%. These results are again in line with the selection channel since more educated trainees have better outside options and are less likely to value the placement jobs than the others. Finally, we do not find any significant effect of our treatment across different caste groups.

Our paper contributes to the relatively thin literature on the effectiveness of vocational training programs in developing countries. The two review papers by Blattman and Ralston (2017) and McKenzie (2017) suggest that vocational training often has limited effect on employment outcomes, although experiments in India by Maitra and Mani (2017) and in Uganda by Alfonsi et al. (2020) show that vocational training can have large positive long-term effects on employment and earnings. A recent follow-up paper by (Bandiera et al., 2022) on the Ugandan experiment highlights the importance of trainees' expectations for the effectiveness of vocational training. Other evidence from India suggests that vocational training provided by Industrial Training Institutes is usually of poor quality (Gasskov et al., 2003; Bertrand and Crepon, 2015). On the DDU-GKY program itself, Chakravorty and Bedi (2019) find that 2-6 months after training completion, the employment rates are not significantly different between DDU-GKY participants and non-participants. Prillaman et al. (2017) also document low rates of employment among DDU-GKY trainees nine months after training. Our contribution is to show that a simple information intervention can make vocational training programs more effective.

Our paper is one of few papers that evaluate information interventions in vocational training programs. Hicks et al. (2011) inform prospective vocational trainees about returns to vocational training, with little effect on enrolment, apart from more females enrolling into male-dominated courses. Jensen (2012) finds that providing information about new work opportunities in call centers to female youth in India increases their demand for vocational training and employment. In a very similar context to ours, Banerjee and Chiplunkar (2018) study the mismatch between DDU-GKY preferences of the trainees and the jobs they are placed in. They find that an intervention which informs placement officers about trainee preferences improves the match between trainees and jobs, and that trainees who were matched with their preferred job stay longer in that job. Our contribution is to show that informing vocational trainees about placement opportunities can improve their placement outcomes in another way: by inducing self-selection of trainees who are a better fit for the available jobs.

We also contribute to the literature on barriers to youth employment in developing countries. The literature emphasises search costs (Franklin, 2018; Abebe et al., 2021b), skills signalling (Carranza et al., 2022; Abebe et al., 2021a; Bassi and Nansamba, 2022; Abel et al., 2020; Groh et al., 2015), or the mismatch between employers' and workers' expectations (Abebe et al., 2017). Our contribution is to show that providing more accurate information about the characteristics of the jobs available to young workers may help them to successfully complete their training and transition to employment.

The paper is structured as follows: Section 2 describes our experimental design and data, Section 3 provides a theoretical framework to interpret the effects of the intervention, Section 4 presents the empirical results and Section 5 concludes.

2 The training program and the experimental design

2.1 The vocational training program

India's DDU-GKY is one of the largest government-sponsored vocational training and placement program in the world, with 1.3 million trainees since its inception in 2014. It targets unemployed rural youth aged 15-35 years from poor families with some secondary education (10th or 12th grade).There are quotas for female and low caste (Scheduled Tribes and Scheduled Castes) trainees.

DDU-GKY relies on a complex institutional set-up. The Ministry of Rural Development (MoRD) and the National Mission Management Unit (NMMU) are responsible for framing policy, and monitoring the scheme. The bulk of the funding, about 75% comes from the central government through the MoRD and the remainder from state governments.DDU-GKY courses are offered based on skills gap assessment studies carried out by the National Skill Development Corporation (NSDC). State Management Missions (SMM) also called State Rural Livelihood Missions (SRLM) are responsible for planning and implementing the program. They invite tenders from private Project Implementing Agencies (PIA), which are then responsible for identifying prospective applicants, providing information on the training courses, delivering training and placing the trained graduates.

The program is mainly residential, and provides a mix of classroom and on-the-job training. Each course consists of two broad components. The first component includes training on soft skills, English and information technology and the second component deals with sector specific training. Depending on the course, the duration of training may be for 3 (576 h), 6 (1152 h), 9 (1578 h) or 12 months (2304 h). The scheme provides for on-the-job training (OJT) for a maximum duration which ranges from 30 days for a 3-month course to 120 days for a 1-year course. The training courses offered by the PIA have to be approved by the National Council for Vocational Training (NCVT) or Sector Skill Councils (SSCs). These Technical Support Agencies (TSA) also provide support in terms of designing the curriculum and certifying the trained graduates.

After the training, PIAs are required to place a minimum of 70% of trainees in jobs which offer regular monthly wages at or above a minimum monthly wage of Rs. 6000. [2] Trainees

[2]Prior to project approval, PIA are required to submit a tentative list of employers to the DDU-GKY administration. This list is part of the PIA's proposal, but the jobs actually provided to candidates may or may not be the same as those on the proposed list. In practice, the placement officer of the PIA liaises with potential employers using all possible networks on a continuous basis. Post-placement, proof of regular wage has to be demonstrated either by a salary slip from the human resource department of the organization or in the absence of a human resource department, a certificate issued by the employer indicating wages paid and

are offered semi-skilled blue collar jobs mostly located in urban sectors. To encourage them to take up and stay in the placement jobs, the scheme has provisions for post-placement financial support.[3] Proof of employment is to be regularly submitted by the PIA to the DDU-GKY administration to avail the post placement financial support. A trained candidate is considered 'employed' only if he/she continues in a PIA job for at least three months. To enhance employment sustainability, PIA are mandated to track all trained/placed candidates for 1 year. During this year, they are also entitled to counseling and guidance.

The study is located in the states of Bihar and Jharkhand, two of the poorest states of India, where job opportunities are scarce, so that most DDU-GKY placement jobs are located in other states (Online Appendix Table B1).

2.2 Intervention

The sample includes 86 batches from training centers located in Bihar and Jharkhand. A batch is a group of students who enrol, have classes, and graduate together. There were 2,488 trainees in total or an average of 30 trainees per batch. The randomization was carried out at the training batch level, stratified by state and sector of training, forming 13 randomization strata. 42 batches were treated (Online Appendix Table B2). All sampled batches were residential programs consisting of 107 days of classroom training on average (between 58 and 205 days) and 17 days of on-the-job-training (between 0 and 60 days).

The intervention was delivered in two classroom sessions (A and B):

- Session A took place in the first two weeks after batch start, before "batch freezing', the time after which no new trainees can be enrolled. Treatment batches were provided with a list of detailed characteristics of *potentially* available placement jobs. Each list was specific to a training-center and trade, and included: job title, company name, location (city and state), and compensation package (net monthly wage and in-kind benefits). The session ends by a Q&A with placement officers.
- Session B took place approximately 10 days before the completion of the classroom training. Trainees from the treatment batches were provided with a detailed list of positions that were *actually* available to them, with job title, company name, location, and compensation package. Trainees were warned about the need to prepare to take up a for possible migration. The session ends by a Q&A with placement officers.

counter signed by the employee along with a bank statement.

[3] An amount of Rs. 1000 per month is available for 2 months in case the placement is within the district of residence; Rs. 1000 per month for 3 months if placement is outside the district but within the state of residence; and Rs. 1000 per month for 6 months if placement is outside the state of residence.

2.3 Data

Our research is based on primary data collected from four rounds of surveys of trainees: the baseline and the midline surveys were conducted face-to-face, and the two endline surveys on the phone (Online Appendix Figure B2). Trainees who were not surveyed at baseline (either because they were absent on the day of the baseline survey, or due to some other reason), were not surveyed in the followup rounds.

- The baseline survey was administered from December 2018 to October 2019, to all (2,488) participants present before batch freezing. We collected information on a wide range of socioeconomic characteristics of the trainee and household, a range of psychometric tests (GRIT, BIG 5, Attitude and self-esteem, life goals, risk preference), expectations, preferences, opportunity cost, and program awareness (Online Appendix Table B5).

- The midline survey was conducted at the end of the classroom training but before the trainees left for their placement jobs. This survey mainly captured the change in expectations of the trainees. Interviews were carried out from March 2019 to January 2020 and covered 1,812 trainees who were present in the training center on the day of the survey (Panel B and C of Online Appendix Table B6).

- The first endline survey was conducted approximately two months after the end of the training, and the second endline five months after the end of the training. We collected information on post-training outcomes focusing on training completion, job placement, and job tenure. The first one took place from May 2019 to April 2020 and cover 2,389 respondents. The second one from August 2019 to May 2020 and covered 2,367 trainees (Panel A of Online Appendix Table B6).

Sample restrictions The Covid-19 pandemic and the lockdown that started on March 24 2020 caused severe disruption to the collection of our endline surveys. To accommodate the disruption, we amended the original focus of the second-endline questionnaire regarding respondents' current status, to ask about status at the time of the 2020 Holi festival (which started on March 9 2020) in order to better anchor the recollection of their activities. Online Appendix Table B3 shows the number of individuals surveyed during the three sub-periods: (i) pre-Holi (before March 9), (ii) between Holi and March 25 2020, and (iii) after March 25 2020. We restrict our analysis to the 2,163 individuals who had their first endline survey before Holi.[4]

[4]We were not able to match 9 observations – see Row 1 Column 4 of online Appendix Table B3.

Attrition The attrition rate for each wave of the surveys, and the p-values associated with the test of no difference across the treated and the control groups, are provided in Online Appendix Table B4. Attrition is very low for the two endline surveys: 4% for the first endline and 5% for the second endline. Attrition in the midline survey is higher (27%) as the survey was only administered to trainees who were present at the time of the interview. Importantly, attrition rates in all survey rounds are similar across the treatment and control groups. The COVID-related sample restrictions is also uncorrelated with treatment assignment (Row 4 of Table B4).

2.4 Summary statistics and balance tests

The full set of variables and their definitions are provided in Online Appendix Tables B5 and B6. Summary statistics of our baseline variables, and the results of the balance tests for randomization, are provided in Online Appendix Table B7. The average age in our sample is 20, and most trainees have some secondary education.[5] There are more female than male trainees, which is a remarkable achievement given the low labor force participation of women in this context generally. In terms of caste, 15% of the trainees are Scheduled Tribes, and 30% Scheduled Castes, which shows that DDU-GKY successfully targets disadvantaged youth.[6] Another evidence of the pro-poor targeting of DDU-GKY is the very high fraction (79%) of trainees from households below the poverty line. Median household earnings are about 9,000 INR (122 USD) a month. Balancing tests suggest that there are no issues with the randomization (see Table B7).

3 Model

We provide a simple theoretical framework to guide the interpretation of our results. It illustrates two potential effects of information on employment and training outcomes: a selection effect, and a job readiness effect. The proofs of the propositions listed below are in Appendix A.1 - A.4.

[5]92% of the sample completed 10th standard, and 56% completed 12th standard. Since having some secondary education is mandated under the DDU-GKY scheme, the trainees have higher levels of education compared to the population parameters. As per PLHS 2017-18, 24 to 28% of the population have secondary or higher qualification in Jharkhand and Bihar respectively.

[6]The constitution of India classifies its citizens into four categories for affirmative action: Scheduled Castes (SCs) and Scheduled Tribes (STs) are various officially designated groups of historically disadvantaged people in India. Other Backward Class (OBC) includes other groups/communities that are not SC or ST, and the rest of the population is classified as general caste. As per Census 2011, the national average of SC population is 16.6% and that of ST population is 8.6%.

Setup and propositions At the time trainees enter the placement job, they compare their reservation utility to the actual utility they derive from the job. R denotes the difference between the reservation utility and the actual value of the job. If $R > 0$, youths leave the job. If $R < 0$ they stay in the job. At the time trainees join the program, they form an expectation about the utility of the job they will be offered at the end of the training. Let V_0 denote the difference between the expected value of the job at the time of joining the program and the actual value of the job. At the time trainees finish the training and before they start in the job, they refine their expectations. Let V_1 denote the difference between the expected value of the job at the time of completing the training and the actual value of the job.

We assume that the learning process is such that $V_1 = \lambda V_0 + \varepsilon$, where ε is a noise parameter centerd around zero that affects the value update during the training, and $\lambda \in [0, 1]$ is the stickiness parameter. Let λ_T denote the parameter for the treated group, and λ_C for the control group. We assume $\lambda_T \leq \lambda_C$: treated individuals learn faster than control ones. Let Z be a binary treatment assignment indicator, which is randomized. Let D be a binary indicator for training completion and placement. Let S be a binary indicator for the individual staying in the DDU-GKY job for at least 5 months.

To sum up, the timing is as follows:

- $t = 0$: individual enters the program if and only if $V_0 > R$.
- $t = 1$: individual completes the training and takes up the job ($D = 1$) iff $V_1 > R$.
- $t = 2$: individual in placement job learns about its true value and decide to stay in the job for at least five months ($S = 1$) iff $R < 0$.

Proposition 1. *The treatment effect on training completion depends on youth expectations:*

$$P(D = 1|Z = 1, V_0 > 0) - P(D = 1|Z = 0, V_0 > 0) < 0,$$
$$P(D = 1|Z = 1, V_0 < 0) - P(D = 1|Z = 0, V_0 < 0) > 0.$$

The information session brings trainees' expectations closer to the true value of the job for them. Hence the treatment discourages over-optimistic ($V_0 > 0$) trainees, who become more likely to drop out of the training and refuse placement, and encourages over-pessimistic ($V_0 < 0$) ones, who become more likely to complete training and accept placement. Online Appendix A.5 Figure A1 illustrates this proposition using a numerical simulation.

Proposition 2. *The treatment increases the probability to stay in job conditional on placement:*

$$P(S = 1|Z = 1, V_1 > R) - P(S = 1|Z = 0, V_1 > R) > 0.$$

Trainees who are placed in the treatment group have more accurate expectations about the placement job, and hence are more likely to stay in the job than in the control.

Proposition 3. *The treatment has an ambiguous effect on the (unconditional) probability of being in the job five months after training:*

- *For trainees for whom the job has a lower value than the outside option ($R > 0$), the treatment does not affect the probability to be in the job, five months after training.*
- *For trainees for whom the job has a higher value than the outside option ($R < 0$) and who are over-optimistic ($V_0 > 0$), the treatment decreases the probability to be in the job five months after training, by decreasing their probability to be placed.*
- *For trainees for whom the job has a higher value than the outside option ($R < 0$) and who are over-pessimistic ($V_0 < 0$), the treatment increases the probability to be in the job five months after training by increasing their probability to be placed.*

Finally, by increasing awareness about the jobs early on in the training, the intervention may help trainees prepare themselves to the transition to employment. We model this as an increase in the true value of the job by τ, leaving the outside option unchanged.

Proposition 4. *For all trainees, an increase in τ will increase training completion, placement and the probability of being in the job conditional on being placed.*

Discussion The model simplifies the decision-making process in two important ways. First, the number of periods and hence the possibilities to drop out is kept to a minimum in the model. In reality, trainees can drop out any time during training and employment spells (e.g., after batch freezing but before midline, after placement but before training completion, after placement but before three months). We refrain from exploiting this variation for simplicity and to preserve statistical power.

Second, in our framework, training completion and job placement are a single decision. In practice the two steps are distinct, and the training may have a value for trainees independently from placement. We choose not to explore this aspect for two reasons: (i) training completion and placement are both posterior to the intervention; and (ii) since the information provided by the intervention focused on placement jobs, it is unlikely to have major effects on training decisions unrelated to placement.

4 Results

4.1 Empirical framework

We restrict our estimation sample to trainees present at baseline. A batch b is in the treatment group if $Z_b = 1$, in the control group if $Z_b = 0$. An individual i in batch $b(i)$, assigned to a randomization stratum $s(i)$, has a vector of baseline characteristics $\mathbf{X_i}$ (control variables). We assume the following partially linear model:

$$\begin{aligned} y_i &= \beta Z_{b(i)} + f(\mathbf{X_i}, s(i)) + \varepsilon_i, \quad \mathbb{E}(\varepsilon|Z, X, s) = 0 \\ Z_{b(i)} &= g(\mathbf{X_i}, s(i)) + u_i, \qquad \mathbb{E}(u|X, s) = 0 \end{aligned}$$

β is the intention-to-treat effect, the parameter of interest in our setting, and $f(.)$ and $g(.)$ are unknown flexible functions. We estimate β using the DoubleML procedure (Chernozhukov et al., 2018; Bach et al., 2021), using random forests to approach $f(.)$ and $g(.)$ (Wright and Ziegler, 2017). We cluster standard errors at the batch level, and compute q-value following the False Discovery Rate method by Benjamini and Hochberg (1995) to handle multiple hypothesis testing.

4.2 Main outcomes

Table 1 presents the results for our main outcomes in columns numbered [1]-[4]. We first consider the probability that the trainee is in the DDU-GKY job five months after training completion (column [1]). This is the unconditional probability based on the full sample: the dependent variable takes the value of 0 for trainees who did not complete the training and those who completed the training but were not placed. In the control group, 33% of all trainees who started the training are in a placement job. This probability is 2.8 percentage points (ppt) (8%) higher in the treatment group. However, the difference is not statistically significant.

The probability of dropping out of the program is 14% in the control group, and not significantly different in the treatment group (column [2]). The probability of being placed among those who completed the training is 50% and is not different in the two groups (column [3]). Column [4] presents the treatment effect on the conditional probability of being in the job for at least 5 months conditional on placement. This probability is estimated to be 11 ppt higher in the treatment group compared to 62% in the control group, a 17% increase. The effect is positive and significant at the 5% level, but the q-value is 12% after adjusting for multiple hypothesis testing.

In summary, the intervention did not affect the dropout probability or the probability of placement conditional on dropout, but we find suggestive evidence that it improved the conditional probability of staying in the DDU-GKY job. Within the framework of our theoretical model, these findings are consistent with selection effects canceling out on average. For example, the increase in dropout among trainees who are a poor fit for the job and a decrease in dropout among trainees who are a good fit for the job, canceling out.

Table 2 reports the results for the additional outcomes collected from the endline surveys (Panel A & B): whether trainees work in the formal sector (column [1]), whether they live outside of their state of origin (column [2]), whether they use skills from training in their current employment (if they are employed; column [3]), and their life satisfaction (column [4]). We do not find any evidence in support that the intervention increased formal employment among trainees: although the estimated treatment effect is about 7% of the control mean (Panel B), it is insignificant. This suggests that some of the positive treatment effects on the probability of being in the placement job after five months were compensated by trainees in the control finding other jobs in the formal sector. We find no significant difference between the two groups, except for an 11% increase in life satisfaction five months after training, which is not robust to adjustments for multiple hypothesis testing.

4.3 Heterogeneity

Tables 3 and 4 report results for the main outcomes by sub-samples defined by gender (women vs. men), caste (Schedules Caste/Scheduled Tribe vs. OBC/General Caste), education (below 12th grade vs. 12th grade and above), and expected salary in the placement job at baseline (distinguished by whether the expected salary is above or below the median of the realized placement salary).[7] Caste and gender correspond to two dimensions of interest for the DDU-GKY program. Education and expectations are natural dimensions of heterogeneity according to our conceptual framework: more educated trainees may be less likely to join the placement job as they have higher outside options, and trainees who expect the placement jobs to pay more than it does may be more likely to be disappointed when they are placed.

We first consider the treatment effects for women and men separately – Table 3 Panels A-B. In the absence of any intervention, women are more likely to be placed in DDU-GKY jobs (60% vs. 37%) and to be working in that job five months after training (74% vs. 39%). This may be due to the fact that DDU-GKY offers rare work opportunities for rural women, whose labor force participation is low (Chatterjee et al., 2015). The intervention

[7] The median is computed within strata defined by state $\times$ trade.

has differential effects by gender: small and insignificant for women, much stronger among men. The effects for men are significant at conventional significance levels and are very large in economic terms: the intervention increases the probability of staying in DDU-GKY jobs conditional on placement by 71%, from 39% to 67%, closing 80% of the gender gap. A possible explanation is that the mismatch between trainees' expectations and the placement job is more of an issue for men because they have a broader range of outside options than women.

We next explore heterogeneity in treatment effects along the caste dimension. On average, trainees from disadvantaged background (SC/ST) are more likely to be placed (57% vs. 44%) and to stay in the placement job (67% vs. 58%) after five months compared to those from OBC/General castes (Table 3 Panels C-D). The heterogeneity of the treatment effect tends to exacerbate the differences across castes: the probability of staying in the placement job conditional on being placed increases by 13 percentage points for SC/ST trainees (p-value of 0.03, q-value of 0.25), compared to 6 percentage points for higher-caste ones (insignificant).

In Panels A-B of Table 4, we study treatment heterogeneity by educational attainment, whether trainees have completed 12th grade or higher or not. As expected, in the control group, conditional on training completion, less educated trainees were more likely to be placed and stay in the job than more educated ones, which is consistent with the fact that the jobs available were only semi-skilled. At the same time, however, less educated trainees drop out twice as often as more educated ones (18% vs. 10%), which may be due to learning difficulties in the training program. Interestingly, the intervention reduces dropout for the less educated by 35% (p-value 0.20) and increases dropout for the more educated trainees by 50% (p-value 0.052). This suggests that better information improves the fit between trainees and jobs. The intervention did not affect placement rates for the two groups, conditional on training completion. Still, it did increase the probability of staying in a DDU-GKY job, conditional on placement for both groups. However, the effect is stronger and significant only for the less educated trainees.

Finally, Panels C-D of Table 4 display the heterogeneity in terms of expected placement salary at baseline. The intervention has a strong and significant positive impact on the probability of staying in placement jobs, conditional on being placed, for trainees with initially low salary expectations. By contrast, the effect for trainees with initially high salary expectations is smaller and insignificant. However, we do not find that the intervention led to higher dropout among high-expectations trainees or lower dropout among low-expectations trainees. This lack of impact is not consistent with our framework: we believe it has to do with the fact that our survey may have failed to extract credible information about expectations. In Appendix Table B8, we present the results of a horse-race between the two variables (education and salary expectation) used in the heterogeneity analysis, and show that education is significantly correlated with future wages and formal employment, while our measure of expectations is not.[8]

4.4 Mechanisms

Following our theoretical framework, our intervention can affect the outcomes via two distinct channels: increasing job readiness and improving selection to remain in training. The selection channel comes from the intervention delivering information about jobs. Better-informed trainees make more time-consistent decisions about completing the training and accepting the placement job; those who get placed are a better fit for the jobs available. The job readiness channel comes from the fact that the intervention prepares trainees better for the transition to employment so that they are more likely to stay in the job once placed.

The heterogeneity results point to an increase in the probability of dropping out of training among more educated trainees and a decrease (if anything) in dropout among the less educated. At the same time, less educated trainees are more likely to stay in the DDU-GKY job once placed following the intervention. This is consistent with the selection mechanism: more educated trainees have better outside options than the jobs offered to DDU-GKY trainees and would be less likely to stay in the DDU-GKY job anyway. If all results were driven by job readiness, we would not have expected an increase in dropout among the more educated. As discussed above, we would have expected the same heterogeneity to occur depending on to baseline salary expectations, which is not the case, and could due to the imperfection of our expectation measure.

We can also explore the effect of the intervention on trainees' expectations about their labor market prospects, using information collected in the midline survey, which was carried out right before the second part of the intervention, before the end of the training. An important caveat is that 27% of trainees were absent at the time of the midline survey, including those who dropped out following the first intervention. This implies that the information collected in this survey is unlikely to capture the mechanisms highlighted in the model. In particular, if the intervention lowered expectations of over-optimistic trainees and induced them to drop out or increased expectations of over-pessimistic trainees and made them stay, average expectations in the trainees still enrolled may not change. With

[8]We still show the results on the heterogeneity by salary expectation for the sake of transparency, and because we committed to in the pre-analysis plan.

this caveat in mind, Table 5 presents the estimated treatment effects on the expectations of the trainees. In Panel A of Table 5, we do not find any significant effect on: (i) the perceived probability of getting a job; (ii) the average wage they expected from this offer; (iii) the range in which they expected this offer to be; (iv) the location of the job. Panel B of Table 5 presents some evidence that trainees in the treatment group revised downwards their willingness to accept a job inside of their state of residence (p-value 0.089) and the likelihood that they would stay 12 months in the state (p-value 0.087). While these effects are small and borderline significant, given the actual placement rates, which are much lower (50% conditional on training completion), they suggest that on average trainees became more realistic about their placement outcomes.

An alternative explanation for our findings, not included in our model, is that the intervention improved the match between trainees' preferences and the jobs offered to them. For example, a better awareness of job available after DDU-GKY may lead trainees to express their preferences to placement officers and choose their preferred options. Banerjee and Chiplunkar (2018) show that providing information to placement officers about trainees' preferences leads to more durable matches. To investigate this possibility, we decompose the placement process into three steps: job offer, offer acceptance, and job placement, and estimate the treatment effect on each of them using the sample of trainees who completed the training. Appendix Table B9 presents the results. There is no evidence that the treatment increases the likelihood of a job offer (column [1]) or the likelihood that the offer is accepted conditional on having been made (column [2]). In contrast, there is a positive but insignificant effect on trainees' likelihood of staying on the job for two months (p-value 0.145). Conditional on staying for two months, they stay for at least five months (p-value 0.033). These results confirm that the treatment improved the fit between trainees and placement jobs by changing the pool of trainees, not by changing the likelihood or the quality of the offers made to them.

5 Conclusion

We conducted a randomized experiment to evaluate an intervention that provided detailed information about placement jobs to trainees of the Indian vocational training program DDU-GKY. We find that better informed trainees were 18% more likely to stay in the jobs they were placed in, with higher effects for lower-caste, less-educated, low-expectations male trainees. We analyze our results through the lens of a conceptual model in which two channels could drive the results: (i) the intervention prepares trainees for DDU-GKY jobs,

(ii) the intervention allows trainees with better outside options to drop out earlier from the training. While we cannot rule out that job readiness does not play a role in this setting, we find suggestive evidence in favor of the self-selection channel.

Our results suggest that providing detailed information about post-training job opportunities can help trainees form more accurate expectations, improve self-selection into training, and improve placement outcomes. Given the low cost and the simplicity of the information sessions, the intervention can easily be scaled up to help the program meet its objectives. Importantly, this kind of intervention should take place early enough in the training spell (or even right before it starts) to minimize costs for trainees and training institutions.

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., 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*, The World Bank.

Abel, M., R. Burger, and P. Piraino (2020): "The value of reference letters: Experimental Evidence from South Africa," *American Economic Journal: Applied Economics*, 12, 40–71.

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.

Bach, P., V. Chernozhukov, M. S. Kurz, and M. Spindler (2021): "DoubleML–An Object-Oriented Implementation of Double Machine Learning in R," *arXiv preprint arXiv:2103.09603*.

Bandiera, O., V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2022): "The Search for Good Jobs: Evidence from a Six-Labor Market Experiment in Uganda," .

Banerjee, A. and G. Chiplunkar (2018): "How Important Are Matching Frictions in the Labor Market? Experimental & Non-Experimental Evidence from a Large Indian Firm," Working paper.

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

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.

Bertrand, M. and B. Crepon (2015): "Improving Youth Labor Market Outcomes in Tamil Nadu Through Skilling: A Firm-Level Survey," J-PAL Policy Research Project for the Governemnt of Tamil Nadu. Final Report.

Betcherman, G. (2004): "Impacts of active labor market programs: New evidence from evaluations with particular attention to developing and transition countries," Working paper.

Blattman, C. and L. Ralston (2017): "Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs," SSRN Discussion Paper 2622220.

Card, D., P. Ibarrarán, F. Regalia, D. Rosas-Shady, and Y. Soares (2011): "The labor market impacts of youth training in the Dominican Republic," *Journal of Labor Economics*, 29, 267–300.

Carranza, E., R. Garlick, K. Orkin, and N. Rankin (2022): "Job Search and Hiring with Limited Information about Workseekers' Skills," *American Economic Review*, 112, 3547–83.

Chakravorty, B. and A. S. Bedi (2019): "Skills Training and Employment Outcomes in Rural Bihar," *The Indian Journal of Labour Economics*, 62, 173–199.

Chatterjee, U., R. Murgai, and M. G. Rama (2015): "Job opportunities along the rural-urban gradation and female labor force participation in India," Policy Research Working Paper Series 7412, The World Bank.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21, C1–C68.

Cho, Y., D. Kalomba, A. M. Mobarak, and V. Orozco (2013): *Gender differences in the effects of vocational training: Constraints on women and drop-out behavior*, The World Bank.

Franklin, S. (2018): "Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies," *Economic Journal*, 128, 2353–2379.

Gasskov, V., A. Aggarwal, A. Grover, A. Kumar, and Q. L. Juneja (2003): "Industrial Training Institutes of India: The efficiency study report," ILO, Geneva, Switzerland.

Groh, M., D. McKenzie, N. Shammout, and T. Vishwanath (2015): "Testing the importance of search frictions and matching through a randomized experiment in Jordan," *IZA Journal of Labor Economics*, 4, 7.

Heckman, J., N. Hohmann, J. Smith, and M. Khoo (2000): "Substitution and dropout bias in social experiments: A study of an influential social experiment," *The Quarterly Journal of Economics*, 115, 651–694.

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.

Hirshleifer, S., D. McKenzie, R. Almeida, and C. Ridao-Cano (2016): "The impact of vocational training for the unemployed: experimental evidence from Turkey," *The Economic Journal*, 126, 2115–2146.

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.

Maitra, P. and S. Mani (2017): "Learning and earning: Evidence from a randomized evaluation in India," *Labour Economics*, 45, 116–130.

McKenzie, D. (2017): "How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence," *World Bank Research Observer*, 32, 127–154.

Prillaman, S. A., R. Pande, V. Singh, and C. T. Moore (2017): "What Constrains Young Indian Women's Labor Force Participation? Evidence from a Survey of Vocational Trainees," Working paper.

Wright, M. N. and A. Ziegler (2017): "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R," *Journal of Statistical Software*, 77, 1–17.

Table 1: Results: Main Outcomes

	In Placement Job after 5m (unconditional)	Training Dropout	Job Placement (conditional)	In Placement Job after 5m (conditional)
	[1]	[2]	[3]	[4]
Treatment	0.028	0.008	0.012	0.109
	(0.043)	(0.019)	(0.053)	(0.051)
p-value	0.515	0.692	0.817	0.032
q-value (MHT)	0.817	0.817	0.817	0.126
Observations	2,070	2,089	1,799	890
Control Mean	0.330	0.136	0.493	0.624
Sample	All	All	Trained	Placed

Notes: See Appendix Table B6 for variable definitions. The dependent variables are all binary indicators taking the value of 1 as follows. Column [1]: The trainee was still in a DDU-GKY job after five months (unconditional); Column [2]: The trainee dropped out of training; Column [3]: The trainee was placed in DDU-GKY job conditional on training completion; Column [4]: The trainee was still in a DDU-GKY job after five months conditional on training completion and placement. All regressions control for baseline characteristics chosen by a random forest approach (Wright and Ziegler, 2017) as well as strata fixed effects. Standard errors account for clustering at the batch level. The reported p-value is for the test of no treatment effect, and the q-value is the p-value of the same test accounting for multiple hypothesis testing (MHT) following the False Discovery Rate method by Benjamini and Hochberg (1995).

Table 2: Results: Short-Term and Long-Term Outcomes

	Formal Job	Outside State	Use Skills from Training	Life Satisfaction
	[1]	[2]	[3]	[4]
Panel A: Two Months after Training				
Treatment	-0.001	0.035	-0.105	-0.945
	(0.046)	(0.061)	(0.075)	(2.205)
p-value	0.989	0.560	0.161	0.668
q-value (MHT)	0.989	0.764	0.645	0.764
Observations	2,089	2,088	961	2,089
Control Mean	0.461	0.431	0.804	72.811
Panel B: Five Months after Training				
Treatment	0.027	0.037	-0.031	7.705
	(0.040)	(0.047)	(0.070)	(3.782)
p-value	0.498	0.440	0.655	0.042
q-value (MHT)	0.764	0.764	0.764	0.333
Observations	2,070	2,070	864	1,222
Control Mean	0.400	0.344	0.687	70.952

Notes: See Appendix Table B6 for variable definitions. The dependent variables [1], [2] and [3] are binary indicators taking the value of 1, and [4] is a continuous variable ranging from 0% to 100%. Column [1]: The trainee was in formal wage employment; Column [2]: The trainee lived outside their home state; Column [3]: The trainee used the skills learned in training in their current occupation; Column [4]: Life satisfaction of the trainees. All regressions control for baseline characteristics chosen by a random forest approach (Wright and Ziegler, 2017) as well as strata fixed effects. Standard errors account for clustering at the batch level. The reported p-value is for the test of no treatment effect, and the q-value is the p-value of the same test accounting for multiple hypothesis testing (MHT) following the False Discovery Rate method by Benjamini and Hochberg (1995).

Table 3: Heterogeneity of treatment effects by gender & caste

	In Placement Job after 5m (unconditional)	Training Dropout	Job Placement (conditional)	In Placement Job after 5m (conditional)
	[1]	[2]	[3]	[4]
Panel A: Female				
Treatment	0.017	-0.017	-0.031	0.047
	(0.061)	(0.025)	(0.062)	(0.042)
p-value	0.783	0.500	0.618	0.263
q-value (MHT)	0.783	0.678	0.706	0.527
Observations	1,081	1,097	974	547
Control Mean	0.460	0.116	0.602	0.744
Panel B: Male				
Treatment	0.076	0.021	0.110	0.282
	(0.066)	(0.032)	(0.092)	(0.089)
p-value	0.252	0.509	0.231	0.001
q-value (MHT)	0.527	0.678	0.527	0.012
Observations	989	992	825	343
Control Mean	0.185	0.158	0.365	0.390
Panel C: Lower Caste				
Treatment	0.045	-0.008	0.019	0.133
	(0.070)	(0.034)	(0.077)	(0.062)
p-value	0.519	0.822	0.802	0.031
q-value (MHT)	0.822	0.822	0.822	0.251
Observations	888	891	770	432
Control Mean	0.380	0.137	0.567	0.669
Panel B: Higher Caste				
Treatment	0.030	0.014	0.025	0.064
	(0.042)	(0.024)	(0.053)	(0.064)
p-value	0.474	0.574	0.637	0.320
q-value (MHT)	0.822	0.822	0.822	0.822
Observations	1,182	1,198	1,029	458
Control Mean	0.291	0.136	0.437	0.579

Notes: The dependent variables are all binary indicators taking the value of 1 as follows. Column [1]: The trainee was in the DDU-GKY job five months after the end of training; Column [2]: The trainee dropped out of training; Column [3]: The trainee was placed in the DDU-GKY job conditional on training completion; Column [4]: The trainee was still in a DDU-GKY job after five months conditional on training completion and placement. "Lower Caste" is a dummy variable equal to one for Scheduled Tribes and Scheduled Caste, and "Higher Caste" a dummy variable for Other Backward Class and General Castes. All regressions control for baseline characteristics chosen by a random forest approach (Wright and Ziegler, 2017) as well as strata fixed effects. Standard errors account for clustering at the batch level.

Table 4: Heterogeneity of treatment effects by education & baseline expectations

	In Placement Job after 5m (unconditional)	Training Dropout	Job Placement (conditional)	In Placement Job after 5m (conditional)
	[1]	[2]	[3]	[4]
Panel A: Less educated				
Treatment	0.039	-0.039	0.025	0.138
	(0.051)	(0.030)	(0.062)	(0.060)
p-value	0.448	0.199	0.692	0.021
q-value (MHT)	0.716	0.397	0.791	0.168
Observations	890	896	755	428
Control Mean	0.379	0.183	0.573	0.664
Panel B: More educated				
Treatment	0.023	0.045	0.007	0.090
	(0.043)	(0.023)	(0.062)	(0.057)
p-value	0.603	0.052	0.915	0.114
q-value (MHT)	0.791	0.207	0.915	0.303
Observations	1,180	1,193	1,044	462
Control Mean	0.293	0.101	0.437	0.587
Panel C: Lower Salary Expectations				
Treatment	0.014	0.014	-0.008	0.145
	(0.046)	(0.022)	(0.057)	(0.061)
p-value	0.762	0.505	0.885	0.018
q-value (MHT)	0.871	0.808	0.885	0.140
Observations	1,205	1,214	1,055	516
Control Mean	0.333	0.127	0.489	0.601
Panel B: Higher Salary Expectations				
Treatment	0.048	-0.010	0.047	0.096
	(0.050)	(0.027)	(0.062)	(0.061)
p-value	0.341	0.728	0.451	0.115
q-value (MHT)	0.808	0.871	0.808	0.460
Observations	865	875	744	374
Control Mean	0.325	0.149	0.499	0.655

Notes: The dependent variables are all binary indicators taking the value of 1 as follows. Column [1]: The trainee was in the DDU-GKY job five months after the end of training; Column [2]: The trainee dropped out of training; Column [3]: The trainee was placed in the DDU-GKY job conditional on training completion; Column [4]: The trainee was still in a DDU-GKY job after five months conditional on training completion and placement. "Less Educated" denotes trainees with less than 12th grade and "More Educated" trainees with 12th grade and above. "Low Expectations" denotes trainees with baseline salary expectations below the median wage earned by trainees of the same batch after placement. "High Expectations" denotes trainees with expectations above the median wage. All regressions control for baseline characteristics chosen by a random forest approach (Wright and Ziegler, 2017) as well as strata fixed effects. Standard errors account for clustering at the batch level.

Table 5: Results: Additional Outcomes

	Treatment	Standard Error	p-value	q-value (MHT)	Control Mean
	[1]	[2]	[3]	[4]	[5]
Panel A: Intermediary Outcomes					
Expected proba of job offer	0.028	(0.091)	0.761	0.915	9.410
Average expected salary	32	(304)	0.915	0.915	11,334
Expected Max - Min salary	-41	(251)	0.869	0.915	3,693
Expected job out of state	0.041	(0.148)	0.780	0.915	8.857
Panel B: Secondary Outcomes					
Expected earnings (in 12 mths)	365	(558)	0.513	0.836	14,616
Preferred earnings (in 12 mths)	903	(678)	0.183	0.475	18,184
Proba training completion	0.056	(0.038)	0.137	0.475	9.800
Training useful?	0.043	(0.076)	0.569	0.836	9.452
Training satisfaction	0.095	(0.068)	0.161	0.475	9.507
Proba accept job in state	-0.359	(0.212)	0.089	0.475	8.548
Proba stay 12 mths in state	-0.362	(0.212)	0.087	0.475	8.501
Proba accept job out of state	0.086	(0.154)	0.579	0.836	8.730
Proba stay 12 mths out of state	-0.098	(0.164)	0.550	0.836	8.618

Notes: The total number of observations used is 1613. The dependent variables are measured at the midline survey. See Appendix Table B6 for variable definitions. Likelihood variables range from 0% to 100%. Panel A comprises of Likelihood of getting a job at the end of the training; the Expected average salary on the job offered at the end of the training (in rupees); The difference between the maximum and the minimum expected salary on the job offered at the end of the training (in rupees); Likelihood of getting a job outside of the state at the end of the training. Panel B shows the treatment effects on Expected earnings after 12 months; Desired earnings after 12 months; Likelihood of completing the training; The degree to which the training is useful; The degree to which the trainees are satisfied with the training; Likelihood of accepting a job in the state; Likelihood of staying 12 months in a job in the state after accepting it; Likelihood of accepting a job outside the state; Likelihood of staying 12 months in a job outside the state after accepting it. All regressions control for baseline characteristics chosen by a random forest approach (Wright and Ziegler, 2017) as well as strata fixed effects. Standard errors account for clustering at the batch level. The reported p-value is for the test of no treatment effect, and the q-value is the p-value of the same test accounting for multiple hypothesis testing (MHT) following the False Discovery Rate method by Benjamini and Hochberg (1995).

Online Appendices

A Theoretical Appendix

A.1 Proof of proposition 1

Given $(D = 1)$ iff $V_1 > R$

$$\begin{aligned}P(D=1|Z=1) - P(D=1|Z=0) =& P(V_1 > R|Z=1) - P(V_1 > R|Z=0) \\ =& P(\lambda_T V_0 + \varepsilon > R) - P(\lambda_C V_0 + \varepsilon > R)\end{aligned}$$

Since $\lambda_T < \lambda_C$ the sign of this expression depends on the sign of V_0:

$$\begin{aligned}&P(V_1 > R|Z=1, V_0 > 0) - P(V_1 > R|Z=0, V_0 > 0) < 0, \\ &P(V_1 > R|Z=1, V_0 < 0) - P(V_1 > R|Z=0, V_0 < 0) > 0.\end{aligned}$$

□

A.2 Proof of proposition 2

$$\begin{aligned}&P(S=1|Z=1, V_1 > R) - P(S=1|Z=0, V_1 > R) \\ &= P(S=1|\lambda_T V_0 + \varepsilon > R) - P(S=1|\lambda_C V_0 + \varepsilon > R).\end{aligned}$$

Since $\lambda_T < \lambda_C$, the expression is negative. □

A.3 Proof of proposition 3

Proof: We first consider people with a negative value of being in the job ($R > 0$). For them, the probability of staying in the job is equal to zero regardless of the treatment, so that the impact of the treatment is mechanically zero:

$$P(S=1|Z=1, R > 0) = P(S=1|Z=0, R > 0) = 0$$

We next consider people with a positive value of being in the job ($R < 0$). They will be placed if and only if they complete the training:

$$\begin{aligned}&P(S=1|Z=1, R < 0) - P(S=1|Z=0, R < 0) \\ &= P(D=1|Z=1, R < 0) - P(D=1|Z=0, R < 0) \\ &= P(\lambda_T V_0 + \varepsilon > R|R < 0) - P(\lambda_C V_0 + \varepsilon > R|R < 0).\end{aligned}$$

The sign of the above expression depends on the sign of V_0:

- For trainees who are over-optimistic ($V_0 > 0$) and have positive value of being in the job ($R < 0$), the treatment effect on the probability of being in the job is negative:

$$P(\lambda_T V_0 + \varepsilon > R | V_0 > 0, R < 0) - P(\lambda_C V_0 + \varepsilon > R | V_0 > 0, R < 0) < 0.$$

- For trainees who are over-pessimistic ($V_0 < 0$) and have a positive value of being in the job ($R < 0$), the treatment effect on the probability of being in the job is positive:

$$P(\lambda_T V_0 + \varepsilon > R | V_0 < 0, R < 0) - P(\lambda_C V_0 + \varepsilon > R | V_0 < 0, R < 0) > 0.$$

$\square$

A.4 Proof of proposition 4

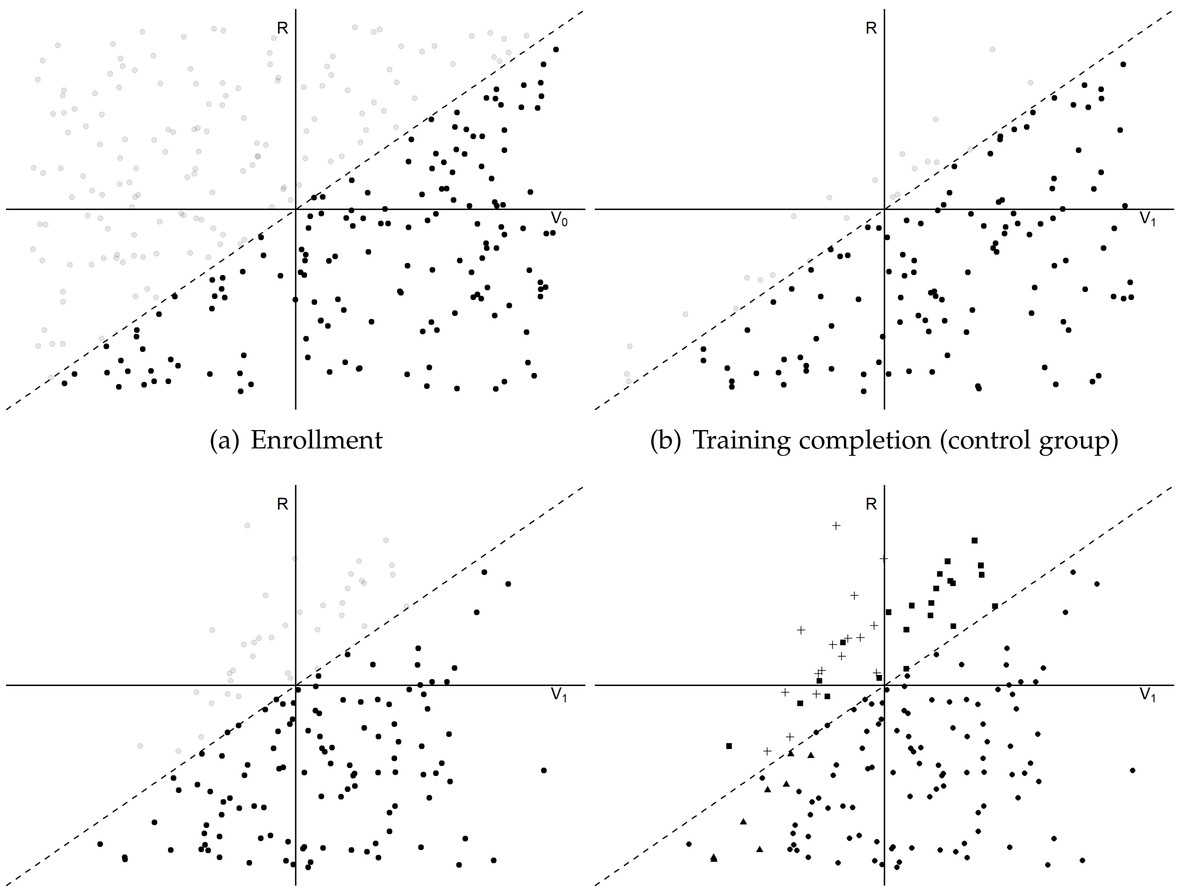
Treated trainees will choose to complete training and get placed iff $\lambda_T V_0 + \tau + \varepsilon > R$, while the criterion remains $\lambda_C V_0 + \varepsilon > R$ for trainees in the control group. Treated trainees who get placed will choose to stay on the job iff $R < \tau$, while their control counterparts will stay iff $R < 0$. $\square$

A.5 Numerical simulation

We illustrate our theoretical predictions using a numerical simulation here. In Panel A, we place individuals as a function of their V_0 and R. Individuals who decide to enrol are all those below the $V_0 = R$ line, for whom the expected value of the job is higher than their outside option.

During the training, individuals receive information about the value of the job. The value of V_1 is shown on the x-axis of Panels B and C, while R remains unchanged on the y-axis. Panel B shows the joint distribution of (V_1, R) in the control group, for whom $\lambda_C = .8$. We see that a few individuals are pushed out of the $V_1 > R$ part, and will choose to drop out. For the treatment group, for whom $\lambda_T = .2$, the distribution of V_1 is much more concentrated around the true value of the job (Panel C). Panel D highlights four types of candidates by comparing the control and treatment distributions. The first group is composed of over-optimistic trainees (square shape), for whom the treatment reduces V_1, which increases their dropout probability. The second group is composed of over-pessimistic trainees (triangle shape), the treatment increase V_1, bringing it closer to the truth, which increases their training completion probability. The third group is composed of trainees who have low value of the training and would drop-out with or without the intervention (cross shape). The fourth group are trainees for whom the training is better than the outside option and who would complete the training with or without the intervention (circle shape).

Figure A1: Theoretical framework: numerical illustration



(a) Enrollment

(b) Training completion (control group)

(c) Training completion (Treatment group)

(d) Training completion (Treatment effect)

B Additional tables and figures

B.1 Background: a description of Bihar and Jharkhand

Table B1: Demographic and economic indicators

Indicators	Bihar	Jharkhand	India
Percentage of total Population of India	8.60%	2.72%	
Total Population	10,38,05,267	3,29,88,134	1,21,08,54,977
Male Population	52.20%	51.32%	51.51%
Female Population	47.80%	48.68%	48.43%
Scheduled Caste Population (SC)	15.96%	12.08%	16.63%
Scheduled Tribe Population (ST)	1.29%	26.21%	8.63%
State Literacy Rate	61%	66%	74%
Male Literacy Rate	69%	77%	82%
Female Literacy Rate	51%	55%	65%
State labor Force Participation Rate (LFPR)	27%	31.40%	
Male LFPR	47.80%	52.20%	55%
Female LFPR	3.70%	9.20%	17%

Source: Census 2011 and NSSO 68th round (2011-12).

B.2 Time of survey waves and attrition

Figure B2: Timing of surveys and interventions in each batch

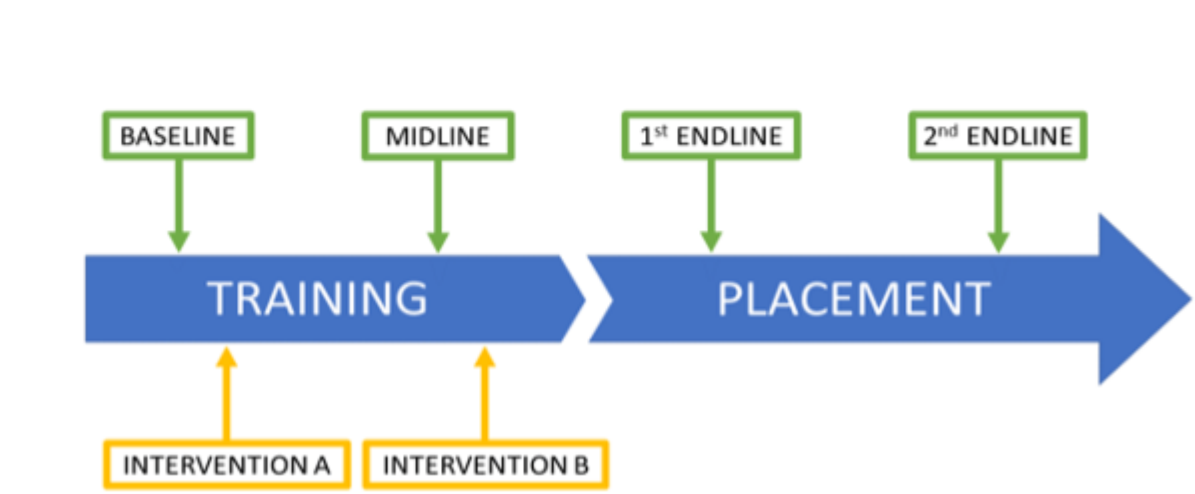


Table B2: Number of batches surveyed by sector and state

Sector	**Control**	**Treatment**	**Total**
BIHAR			
Apparel	3	3	6
Banking, Financial Services, and Insurance	3	3	6
Construction	7	7	14
Healthcare	1	1	2
IT	6	5	11
Logistics	1	1	2
Retail	4	4	8
Tourism-hospitality	3	2	5
TOTAL	28	26	54
JHARKHAND			
Apparel	9	8	17
Automotive	1	1	2
Construction	4	5	9
Healthcare	1	1	2
Security	1	1	2
TOTAL	16	16	32

Table B3: Endline completion dates and sample sizes

			Second Endline		
	Time of survey	Pre Holi	Holi - 25 Mar	Post 25 Mar	Total
	[1]	[2]	[3]	[4]	[5]
First Endline	Pre Holi	1092	187	893	2172
	Holi - 25 Mar	0	0	62	62
	Post 25 Mar	0	0	254	254
	Total	1092	187	1209	2488

Notes: This table shows the time periods for the first and second endlines. Column [1] presents the survey periods, and Columns [2]-[5] present the numbers surveyed during the respective periods. The start of the Holi festival was the 9th March, 2020.

Table B4: Attrition Rates

Survey Timing	Control Mean	Treatment Effect	p-value
	[1]	[2]	[3]
Attrition Midline	0.268	0.004	0.900
Attrition First Endline	0.041	-0.003	0.662
Attrition Second Endline	0.049	-0.001	0.886
COVID-related Sample Restriction	0.094	-0.010	0.777

Notes: Column [2] is obtained from the regression of attrition dummy on an intercept and the treatment indicator, controlling for strata fixed effects. The p-values associated with the test of no effect of treatment, are in Column [3]. The number of observations is 2,488.

B.3 Variable definitions

Table B5: Definition of control variables

Variable Name	Variable Definition
Panel A: Individual Characteristics	
Female	Trainee is female
Older (More than 20)	Age of trainee more than 20 years
Married	Trainee is married
Caste: ST	Caste is Scheduled Tribe
Caste: OBC	Caste is Other Backward Class
Caste: General	Caste is General
Religion: Muslim	Religion is Muslim
Religion: Christian	Religion is Christian
Panel B: Education	
Middle school (6-8 class)	Educated till 6-8 class
Lower secondary (9-10 class)	Educated till 9-10 class
Tertiary education (Graduate & above)	Educated till graduation and above
Matriculation exam (Class X)	Passed class 10th
Exam score more than 50%	Matric exam score more than 50%
Higher secondary exam (Class XII)	Passed class 12th
Exam score less than 50%	12th exam score less than 50%
Panel C: Skills	
Big 5 Extraversion Test (1 to 5)	Set of forty-four questions, scored between "strongly disagree" (1) to "strongly agree" (5)
Big 5 Agreeableness Test (1 to 5)	
Big 5 Conscientiousness Test (1 to 5)	
Big 5 Neuroticism Test (1 to 5)	
Big 5 Openness Test (1 to 5)	
Grit Test (1 to 5)	Set of seven questions, scored between "Very much like me" (1) to "Not much like me at all" (5)
ASE Test (1 to 4)	Set of ten questions, scored between "Strongly agree" (1) to "Strongly Disagree" (4)
Life goal Test(1 to 4)	Set of five questions, scored between "Very Important" (1) to "Not at all important" (4)
Duration of baseline survey (above median)	Duration of baseline survey more than median

Notes: These are based on information captured during the baseline survey.

Table B5: Definition of control variables (continued)

Variable Name	Variable Definition
Panel D: Family Background	
Household head relationship (mother)	Head of household is mother
Household head relationship (others)	Head of household is other than mother and father
Immediate difficulty to family immediate	Difficulty to family if the trainee is in training center during training
Future difficulty to family	Difficulty to family if the trainee is in job outside the state for 12 months
Earning members (3 or more)	Number of earning members in the household
Household earning (15000 or more)	Household earning
Household earning (5000 or less)	
Household earning (5001-9000)	
Agriculture land	Household has agriculture land
BPL card	Household is below poverty line
RSBY card	Household covered under government health insurance scheme
MNREGA	Any member of household worked in MNREGA in the past one year
SHG member	Any member of household is a member of self-help group
Semi pucca house	Type of house
Pucca house(IAY)	
Pucca house(Non IAY)	
Own house	If the household own a house
Internet use	If the trainee had used internet before joining training
Joint household	Type of household
Household members (2 or less)	Number of household members
Household members (6 or more)	
Ever migrated out of state (self)	Trainee migrated out of state in the last one year
Ever migrated out of state (relatives)	Any household member migrated out of state in the last one year
Relatives migrated (one)	Number of household members migrated out of state in the last one year
Relatives migrated (2 or more)	

Note: Difficulty variables are expressed as a fraction between zero and one.

Table B5: Definition of control variables (end)

Variable Name	Variable Definition
Panel E: Expectations	
Previous earning	Previous earning of trainee
Hypothetical earning (immediate)	Hypothetical earning now, had he/she not participated in training
Hypothetical earning (in one year)	Hypothetical earning one year from now, had he/she not participated in training
Expected earning (in one year)	Expected earning one year from now, after training completion
preferred earning (in one year)	Preferred earning one year from now, after training completion
Expected minimum salary (immediate)	Expected minimum salary in job after training
Expected maximum salary (immediate)	Expected maximum salary in job after training
Expected average salary (immediate)	Expected average salary in job after training
Training awareness	Awareness of the training scheme
Training usefulness	Perceived usefulness of the training program
Training satisfaction	Perceived satisfaction with the decision to participate
Likelihood of training completion	Likelihood of completing the training program
Likelihood of job offer	Likelihood of getting a job offer after training completion
Likelihood of job offer outside state	Likelihood that the job offer will be outside the residence state
Likelihood of accepting job inside state	Likelihood of accepting a job if the job is in the residence state
Likelihood of retention in job inside state	Likelihood of staying in this job for 12 months
Likelihood of accepting job outside state	Likelihood of accepting a job if the job is outside the residence state
Likelihood of retention in job outside state	Likelihood of staying in this job for 12 months

Notes: The earning variables are dummy variables equal to one if the individual response is above the median in the stratum (state$\times$trade). The likelihood variables are expressed as a fraction between zero and one.

Table B6: Definition of outcome variables

Variable Name	Variable Definition
Panel A: Main Outcomes	
In placement job after 5 months (unconditional)	In DDU-GKY job after 5 months of training completion
Training dropout	The trainee dropped out of training during training tenure
Job placement (conditional)	Placed in a DDU-GKY job conditional on training completion
In placement job after 5 months (conditional)	Being in DDU-GKY job after 5 months of training completion conditional on placement
Panel B: Intermediary Outcomes	
Likelihood job offer (0-10)	Likelihood of getting a job offer after training
Average salary (Rs)	Expected average salary on the job after training
Difference maximum - minimum salary (Rs)	Difference of maximum and minimum expected salaries in job after training
Likelihood job offer out of state (0-10)	Likelihood of the job being outside the residence state
Panel C: Secondary Outcomes	
Expected Earnings in 12 months	Expected earnings after 12 months of training
Preferred Earnings in 12 months	Preferred earnings after 12 months of training
Likelihood training completion (0-10)	Likelihood of completing the training course
Training usefulness (0-10)	Perception of training usefulness
Training satisfaction (0-10)	Satisfaction with training
Likelihood accept job in state (0-10)	Likelihood of accepting a job if the job is within the residence state
Likelihood stay 12 months in state (0-10)	Likelihood of staying in a job if the job is within the residence state for 12 months
Likelihood accept job out of state (0-10)	Likelihood of accepting a job if the job is outside of the residence state
Likelihood stay 12 months out of state (0-10)	Likelihood of staying in a job if the job is outside of the residence state for 12 months
Panel D: Additional Outcomes	
Formal job	The trainee was in a wage/salaried job
Outside state	The trainee lived outside the residence state
Use skills from training	Used the skills learned in training in their current occupation;
Life satisfaction (0-10)	Life satisfaction of the trainees

Notes: The likelihood variables are scored from 0% (least likley) to 100% (most likely).

B.4 Summary statistics and balance tests

Table B7: Baseline summary statistics (averages) and balance tests - [Part 1 of 3]

Variable	Control Group	Treatment Group	Diff [2]-[1]	p-value
	[1]	[2]	[3]	[4]
Panel A: Demographics and Caste				
Female	0.564	0.585	0.021	0.758
Older (More than 20)	0.280	0.241	-0.039	0.131
Married	0.093	0.112	0.019	0.258
Caste:ST	0.153	0.146	-0.007	0.811
Caste:OBC	0.484	0.524	0.040	0.331
Caste:General	0.066	0.078	0.012	0.436
Religion:Muslim	0.059	0.054	-0.005	0.801
Religion:Christian	0.034	0.035	0.001	0.962
Panel B: Education				
Middle school (6-8 class)	0.071	0.054	-0.017	0.142
Lower secondary (9-10 class)	0.362	0.364	0.002	0.950
Tertiary education (Graduate & above)	0.091	0.084	-0.007	0.670
Matriculation exam (Class X)	0.906	0.924	0.018	0.180
Exam score more than 50%	0.466	0.487	0.021	0.416
Higher secondary exam (Class XII)	0.553	0.565	0.012	0.756
Exam score less than 50%	0.228	0.216	-0.012	0.543
Panel C: Skills				
Big 5 Extraversion Test (1 to 5)	3.294	3.289	-0.005	0.869
Big 5 Agreeableness Test (1 to 5)	3.747	3.775	0.028	0.338
Big 5 Conscientiousness Test (1 to 5)	3.815	3.916	0.101	0.009
Big 5 Neuroticism Test (1 to 5)	2.460	2.420	-0.040	0.333
Big 5 Openness Test (1 to 5)	3.904	4.007	0.103	0.030
Grit Test (1 to 5)	3.374	3.450	0.076	0.039
ASE Test (1 to 4)	2.104	2.084	-0.020	0.352
Life goal Test(1 to 4)	2.147	2.136	-0.011	0.588
Duration of baseline survey (above median)	0.501	0.541	0.040	0.362
Number of observations	1041	935		

Notes: Variable definitions are provided in Online Appendix Tables B6 and B5. Columns [1] and [2] report the mean value in the control group and treatment group respectively. Treatment dummy coefficient estimates in the regression of the variable, controlling for the strata fixed effects are in column [3]. All standard errors account for clustering at the batch level. The p-value associated with the test of no treatment effect is in column [4]. Total number of observations used is 1,976.

Table B7: Baseline summary statistics (averages) and balance test (cont'd) [Part 2 of 3]

Variable	Control Group	Treatment Group	Diff [2]-[1]	p-value
	[1]	[2]	[3]	[4]
Panel D: Socioeconomic background				
Household head relationship (mother)	0.077	0.084	0.007	0.560
Household head relationship (others)	0.084	0.124	0.040	0.009
Immediate difficulty to family	0.092	0.108	0.016	0.296
Future difficulty to family	0.138	0.150	0.012	0.543
Earning members (3 or more)	0.087	0.113	0.026	0.088
Household earning (15000 or more)	0.135	0.186	0.051	0.016
Household earning (5000 or less)	0.314	0.245	-0.069	0.020
Household earning (5001-9000)	0.225	0.228	0.003	0.895
Agriculture land	0.603	0.667	0.064	0.050
BPL card	0.794	0.759	-0.035	0.188
RSBY card	0.371	0.345	-0.026	0.336
MNREGA	0.252	0.209	-0.043	0.062
SHG member	0.764	0.770	0.006	0.813
Semi pucca house	0.202	0.226	0.024	0.310
Pucca house(IAY)	0.099	0.076	-0.023	0.118
Pucca house(Non IAY)	0.191	0.220	0.029	0.272
Own house	0.995	0.994	-0.001	0.662
Internet use	0.478	0.488	0.010	0.793
Joint household	0.062	0.076	0.014	0.239
Household members (2 or less)	0.061	0.045	-0.016	0.117
Household members (6 or more)	0.384	0.379	-0.005	0.853
Ever migrated out of state (self)	0.121	0.121	0.000	0.986
Ever migrated out of state (relatives)	0.498	0.529	0.031	0.296
Relatives migrated (one)	0.364	0.357	-0.007	0.754
Relatives migrated (2 or more)	0.134	0.172	0.038	0.146
Number of observations	1,041	935		

Notes: Difficulty variables are expressed as a fraction between zero and one. Also see notes provided with the first part of this Table [Part 1 of 3].

Table B7: Baseline summary statistics (averages) and balance test (end) [Part 3 of 3]

Variable	Control Group	Treatment Group	Diff [2]-[1]	p-value
	[1]	[2]	[3]	[4]
Panel E: Expectations				
Previous earning	0.099	0.105	0.006	0.712
Hypothetical earning (immediate)	0.113	0.144	0.031	0.148
Hypothetical earning (in one year)	0.184	0.203	0.019	0.557
Expected earning (in one year)	0.379	0.411	0.032	0.484
Preferred earning (in one year)	0.418	0.454	0.036	0.429
Training awareness	0.546	0.528	-0.018	0.369
Training usefulness	0.936	0.931	-0.005	0.510
Training satisfaction	0.947	0.947	0.000	0.934
Likelihood of training completion	0.952	0.944	-0.008	0.229
Likelihood of job offer	0.904	0.898	-0.006	0.523
Expected minimum salary (immediate)	0.372	0.406	0.034	0.546
Expected maximum salary (immediate)	0.384	0.422	0.038	0.444
Expected average salary (immediate)	0.449	0.487	0.038	0.473
Likelihood of job offer outside state	0.787	0.794	0.007	0.672
Likelihood of accepting job inside state	0.848	0.844	-0.004	0.758
Likelihood of retention in job inside state	0.838	0.832	-0.006	0.688
Likelihood of accepting job outside state	0.827	0.828	0.001	0.935
Likelihood of retention in job outside state	0.820	0.818	-0.002	0.884
Number of observations	1041	935		

Notes: Earning variables are dummy variables equal to one if the survey response is above the median in the stratum (state$\times$trade). Likelihood variables are expressed as a fraction between zero and one. Also see notes provided with the first part of this Table [Part 1 of 3].

B.5 Additional results

Table B8: How do salary expectations and education explain wages and formal employment

	Wages (endline 1)		
	(1)	(2)	(3)
Salary expectations: Below median	43		61
	(154)		(153)
Education: Below 12th grade		-440	-442
		(151)	(151)
Adj. R-squared	-0.001	0.006	0.005
Observations	1270	1270	1270
	Formal employment (endline 2)		
	(1)	(2)	(3)
Salary expectations: Below median	0.012		0.009
	(0.021)		(0.021)
Education: Below 12th grade		0.065	0.065
		(0.021)	(0.021)
Adj. R-squared	-0.000	0.004	0.004
Observations	2363	2363	2363

Notes: This table show the results of regressions of wages (measured during the first endline) and being formally employed (measured during the second endline), as a function of the measure of salary expectations and education that we use in the heterogeneity analysis.

Table B9: Results: Job offer, acceptance, placement, and staying in job

	Received a job offer	Accepted job offer cond. on [1]	Placed for 2 months cond. on [2]	Placed for 5 months cond. on [3]
	[1]	[2]	[3]	[4]
Treatment	-0.026	-0.033	0.067	0.116
(standard error)	(0.046)	(0.038)	(0.046)	(0.053)
p-value	0.576	0.390	0.145	0.033
q-value (MHT)	0.577	0.521	0.29	0.134
Observations	1688	1391	1091	883
Control Mean	0.833	0.795	0.790	0.623

Notes: The dependent variables are all binary indicators taking the value of 1 as follows. Column [1]: The trainee received job offer conditional on training completion; Column [2]: The trainee accepted the job conditional on the job offer being made; Column [3]: The trainee was placed in a DDU-GKY job conditional on job acceptance and remained for 2 months; Column [4]: The trainee was still in DDU-GKY job after 5 months conditional on placement. All regressions control for baseline characteristics chosen by a random forest approach (Wright and Ziegler, 2017) as well as strata fixed effects. Standard errors account for clustering at the batch level. The reported p-value is for the test of no treatment effect, and the q-value is the p-value of the same test accounting for multiple hypothesis testing (MHT) following the False Discovery Rate method by Benjamini and Hochberg (1995).