# Can information about jobs improve the effectiveness of vocational training? Experimental evidence from India

Bhaskar Chakravorty

Wiji Arulampalam

Clément Imbert

Roland Rathelot*

January 23, 2021
*Preliminary version. Do not circulate without permission.*

**Abstract**

Vocational training is commonly used to increase youth employment, but its effectiveness is often limited by the mismatch between youth expectations and jobs available to them. We provide information about prospective jobs to randomly chosen batches of DDU-GKY, a large-scale training program in India. We find that better informed trainees are 17% more likely to stay in the jobs in which they are placed. We argue that this effect is driven by improved selection into placement, rather than higher job readiness. The intervention has no effect on training completion on average but increases dropout of trainees with higher education (who have better outside options) and reduces dropout of trainees with lower education (who have worse outside options). Our findings suggest that better information about prospective jobs during vocational training can enhance its effect on youth employment.

# 1 Introduction

Youth unemployment or underemployment, is a major issue 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, and so does Hirshleifer et al. (2016) in Turkey, Card et al. (2011) in Dominican republic, and Cho et al. (2013) in Malawi. One potential reason for this attrition is the mismatch between youth expectations and the jobs available to them (Abebe et al., 2017a).

In India, the context of our study, around 5-7 million young workers enter the labour market every year. A large majority of them are unskilled and come from rural areas. As part of its "Skilling India" strategy, the government launched a national training program DDU-GKY (Deen Dayal Upadhyay Grameen Kaushalya Yojana) in 2014, targeting disadvantaged rural youths. DDU-GKY is a public-private partnership program where the training partners - called Project Implementing Agency (PIA), provide the training to a group of students (known as a batch) at a time, on different trades at their training centres. The program is mainly residential, lasting for about three to six months. According to the official website, the Indian government has invested more than Rs. 56 billion (roughly $ 800 million) on this program. Since its inception, DDU-GKY has trained over a million youths. Although every trainee is offered a job, the placement performance is disappointing for the government. So far, only about 55% of the million trainees who entered the programme have been effectively placed.[1]

In this paper, we test whether the effectiveness of vocational training programmes can be enhanced by providing better information regarding post-training placement opportunities. We implement a randomised experiment across two Indian states (Bihar and Jharkhand). Out of 86 DDU-GKY batches, 42 were randomly selected to receive two information sessions about placement opportunities. The main outcomes of interest of this study are: training completion, successful job placement in a DDU-GKY job on completion, and continuing in the placed-job for at least five months.

We develop a conceptual framework and argue that the lack of information about prospective jobs can reduce the effectiveness of vocational training on employment through two channels. First, the training may attract candidates who have a reservation utility from

---

[1]Official statistics from http://ddugky.gov.in/ on 3 Sept 2020.

a job that is too high compared to the placement jobs provided by the scheme (selection effect). Second, the transition between training and employment may be difficult if the trainees come unprepared (job readiness effect). While better information is expected to increase the probability of staying in the job conditional on placement in either cases, it will have different effects on training completion depending on the channel. On the one hand, better job preparedness increases the value of the job, and should decrease dropout and increase placement. On the other, improved selection into placement will increase dropout among trainees with higher reservation utility and decrease dropout among trainees with lower reservation utility, with ambiguous effects overall.

We find that trainees in the treatment group are 17% more likely to stay in the jobs in which they were placed, which shows that better information about job prospects does improve job retention. There is no effect on average however on dropout or placement, which suggests that the effects may not be driven by better job preparedness. Consistent with the selection effect, once informed about the jobs on offer, we find the youths with higher education are more likely to drop out of training and the youth with lower education are more likely to complete the training.

One of the core objectives of DDU-GKY is to improve employment among the disadvantaged and marginalised population (females, scheduled castes/scheduled tribes (SC-STs), less educated). Hence, we are particularly interested in the impact of our interventions on these individuals who might suffer from larger information barriers on labour market opportunities. The results indicate that for women who already fare better than men in all main outcomes (training completion, placement rates, probability of staying on in the job) in the absence of our intervention, we do not find any significant effect of our intervention. In contrast, we find large effects of the intervention on male trainees. For them, the treatment increases the probability of staying five months in the placement job by as much as 55%. The effects are broadly comparable between along caste lines, although we find slightly larger treatment effects for SC-ST candidates (21% increase in the probability of staying in the placement job).

We provide further evidence on the mechanisms using post-intervention data on trainees expectations. Consistent with negative updating, we find some evidence that on average trainees are less optimistic about their likelihood of taking-up and staying in placement jobs. These results are not fully conclusive, however, since the data were collected a few months after the beginning of the training, and do not provide information on how expectations changed for trainees who dropped out of the training. While our preferred interpretation is that following the intervention, the trainees who selected into the place-

ment stage were a better fit for the jobs available, an alternative explanation to our results could be that the trainees received offers that were a bit fit for them. We test whether the intervention affected offers made to the trainees, and the likelihood that there were accepted, and find no effect.

Our paper contributes to the relatively thin literature on vocational training programs in developing countries. The two review papers by Blattman and Ralston (2017) and McKenzie (2017), suggest that vocational training often have limited effect on employment outcomes, although experiments in Uganda by Alfonsi et al. (2020) and in India by Maitra and Mani (2017), show that vocational training has large positive long-term effects on employment and earnings. 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 programme 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 programmes more effective.

Our paper is one of few papers that evaluate information interventions in vocational training programs. In an experiment in Mongolia, Alzúa et al. (2020) find that informing trainees about the details of the job opportunities post-training significantly increases job duration. Hicks et al. (2011) implement a similar intervention in Kenya. Jensen (2012) finds that providing information on the nature of prospective jobs raises female employment in India. In a paper closely related to our paper, 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 trainees 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 trainees about job characteristics can increase the probability that the trainees stay in jobs in another way: by improving the self-selection of trainees who are a better fit for the jobs available.

We also contribute to the literature on barriers to youth employment in developing countries. The literature emphasises search costs (Franklin, 2018; Abebe et al., 2017b), skills signalling (Abebe et al., 2018; Bassi and Nansamba, 2019; Orkin et al., 2020; Abel et al., 2020; Groh et al., 2015). In a recent paper, Abebe et al. (2017a) highlight that the mismatch between workers and firms expectations may explain the low job finding rate for unemployed youth. 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 a better matched job.

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 briefly concludes.

# 2 The training program and the experimental design

## 2.1 The vocational training programme DDU-GKY

Launched in September 2014, "Deen Dayal Upadhyaya Grameen Kaushal Yojana" (DDU-GKY) is one of India's major and most prominent skills and job creation schemes. This is a national program implemented in all 28 States and Union Territories (DDU-GKY Programme Guidelines, 2016). It focuses on unemployed rural youth aged 15-35 years from poor families, and those out of school. The scheme sponsored by the central and state governments, is monitored at the state level by the state livelihood missions. Currently, over 700 state-approved training agencies (Project Implementing Partners - PIAs) train youths in more than 500 different trades from 50 industrial sectors. Training provided under DDU-GKY, is mainly residential, and is a mix of classroom and on-the-job training. Training and placements are the responsibilities of the state approved private training organisations - PIAs. Most of the placement jobs are located in a state other than the residential state of the candidate, requiring them to migrate far from their family.

## 2.2 Intervention

Two interventions were administered within a cluster randomised trial design with two arms where the treatment groups received both interventions, and none in the control group. The randomisation was carried out at the batch level, stratified by state and sector/trade, forming 13 randomisation strata (Appendix A7, A8). A batch is a group of students who enrol, have classes, and graduate together.

The aim of the study was to test whether an intervention that delivered information about prospective jobs in two sessions (A and B) during the training program, will increase the effectiveness of training, by reducing training dropout, and increasing job uptake and job duration.

- Session A took place in the first two weeks after batch start, before "batch freezing". Training agencies are not allowed to enrol more candidates for that batch after the

batch is 'frozen'. Trainee dropout before batch freezing has low cost for training providers since they have time to attract more students to meet their quantitative objectives. To administer session A, we gathered trainees of the pre-selected treatment batch in a classroom, and provided them with a list of detailed characteristics of *potentially* available placement jobs. The list was training-centre-specific and trade specific. The list included job title, company name, location (city and state), and compensation package (net monthly wage and in-kind benefits). Trainees had the opportunity to ask the placement officer questions regarding the contents of the list.

- Session B took place approximately 10 days before completion of classroom training; before the batch left the training centre campus for placement. The recipients of this intervention were those about to complete their classroom training. We gathered trainees in a classroom and provided them with a list of detailed characteristics of positions that were *actually* available to the candidates of that particular treatment batch for placement. Unlike session A where the job list was trade specific, the job lists were training-centre-specific and batch specific in session B.The list included job title, company name, location (city and state), and compensation package (net monthly wage and in-kind benefits). Trainees had the opportunity to ask the placement officer questions regarding the contents of the list. The intervention also warned trainees about the need to prepare for possible migration.

## 2.3  Sampled batches

In total, from the States of Bihar and Jharkhand, 86 batches were sampled consisting of 42 treatment batches and the rest control batches. The total number of trainees covered was 2488 giving an average of 30 candidates per batch. All batches in the sample had residential mode of training. Average days of classroom training in the sample was little more than 3 months (107 days), ranging from 58 days to 205 days. Average number of days of on-job-training was about 2 weeks (17 days), ranging from 0 days to 60 days.

## 2.4  Data

Our research is based on primary data collected from four rounds of surveys of trainees from the 86 training batches across both Bihar and Jharkhand.

All surveys were administered on tablets using custom designed questionnaires on Survey CTO platform. The data collectors administered the baseline and mid-line surveys in

face-to-face interview sessions with individual trainees, and the two endline surveys using phone interviews. The four survey rounds are described next.

- Baseline survey: the survey was administered to all participants within first two weeks of training inception (i.e before batch freezing which generally occurs after two weeks from the batch start date). The baseline questionnaire is custom designed to capture a wide range of socioeconomic characteristics of the candidate and household, along with a range of psychometric tests (GRIT, BIG 5, Attitude and self-esteem, life goals, risk preference), and expectations, preferences, opportunity cost and program awareness. The baseline survey was conducted from December 2018 till October 2019 covering 2488 respondents.

- Mid-line Survey: this round was conducted at the end of the classroom training and before the batch of trainees left for their placement jobs. All follow-up survey questionnaires were only administered to those respondents who were interviewed at baseline. Candidates 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 follow-up rounds. The mid-line survey mainly captures the change in expectation of the trainees. Interviews were carried out from March 2019 till January 2020 covering 1812 respondents.

- First and second endline surveys: the first endline was conducted approximately two months after training completion, and the second one about five months after training completion. The objective was to collect information about the post-training outcomes focusing on training completion, job placement, and seniority in job. The first endline survey took place from May 2019 to April 2020 and surveyed 2389 respondents; second endline survey was conducted from August 2019 till May 2020, and covered 2367 respondents.

**Attrition**  Table 1 shows the attrition rate for each wave of the surveys, and the p-values associated with the test of no difference across the two groups. The attrition is relatively high at 27% in the midline survey (when some intermediary outcomes are measured), because the midline survey was done in person, and candidates who either had dropped out or were temporarily absent did not take the survey. In contrast, attrition was very low in the two endline surveys: 4% for the first endline and 5% for the second endline. Importantly, as Table 1 shows, the attrition rates in all survey rounds was similar in the treatment and control groups.

**Sample restrictions** The national lockdown due to the Covid-19 pandemic imposed by the Indian Government on the 24th of March 2020, caused major disruption to our data collection for the endline surveys, especially to the remaining second endline surveys. The original second endline questionnaire collected details about the 'current' status of the trainees. There were still some outstanding second endline surveys at the time of the lockdown. Since everything was closed during the lockdown, we changed the focus of the questionnaire to information regarding their status at the time of the annual Holi festival.[2] We felt that the respondents would be able to better recollect what they were doing just before the Holi festival, given the short recall period and the importance of Holi (last second endline was carried out about one and half months from Holi).

Appendix Table A6 provides the numbers of individuals surveyed during the three sub-periods: (i) pre-Holi, (ii) between Holi and until country wide lockdown imposed on 25th March 2020, and (iii) after the 25th March 2020. We use the following two criteria to select the sample of individuals for the analysis involving the two endline survey data. First, we keep only those individuals who had their first endline survey before Holi, since these are unaffected by the Covid shock. This reduced the sample to 2163 observations.[3] Second, of these 2163 observations, we select only those individuals who had their second endline interviews conducted prior to Holi (as they were not effected) and those interviewed during the lockdown (after 25th March 2020) as the modified second endline questionnaire had already been field tested and was in use. We however, dropped the second endline observations from the interviews held between Holi and the start of the lockdown as these observations were used for field testing of the modified second endline questionnaire. Thus, our final analysis sample consists of 1976 observations. It is crucial to note that the restricted sample is not effected by the treatment status. (row 4 of Table 1)

## 2.5 Summary statistics

Summary statistics of our baseline variables, and the results of the balance tests are provided in Table 2. The average age of our participants is 20 and most trainees have some secondary education. The training programme enrolls more female than male candidates, 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, 30% Scheduled Caste and half of the trainees are OBC, which reflects the targeting of the programme towards disadvantaged population is effective. Another evidence of the pro-poor target-

---

[2]Holi festival is a two-day annual event which started on the 9th of March 2020 this year.
[3]We were not able to match 9 observations - see Row 1 Column 4 of Appendix Table A6.

ing of DDU-GKY is the very high fraction (79%) of candidates from households Below the Poverty Line. For more than half the sample, household earnings are less than 9,000 INR (about 122 USD) a month.

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

$$X_i = \beta Z_{b(i)} + \delta_{s(i)} + \varepsilon_i.$$

where $i$ is the trainee, $b(i)$ the batch to which the trainee belongs, $Z_b$ a dummy variable equal to one if batch $b$ is in the treatment group, $\delta_s$ fixed effects for randomisation strata. We test the null of no difference between the treatment and control groups ($\beta = 0$). We use randomisation inference to handle clustering and finite-sample inference issues, and also correct for multiple hypothesis testing, controlling for false discovery rates. Our balance tests suggest that there are no issues with the randomisation.

# 3  Model

This section describes a theoretical framework which will guide the empirical analysis. It illustrates two potential effects of information on employment and training outcomes: a selection and a job readiness effect.

## 3.1  Set-up

Notations:

- Let $R$ denote the difference between the outside option (reservation utility) and the actual value of the job unknown to the trainee at the start of the program. $R > 0$ for youth who would leave the job if they are placed, and $R < 0$ for youth who would stay in the placed job.

- Let $V_0$ denote the difference between the expected value of the job at the time of joining the programme and the actual value of the job. $V_0 > 0$ for over-optimistic youth, $V_0 < 0$ for over-pessimistic youth.

- 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. $V_1 > 0$ for over-optimistic youth, $V_1 < 0$ for over-pessimistic youth.

- The learning process is such that $V_1 = \lambda V_0 + \varepsilon$, where $\varepsilon$ is a noise parameter centred around zero that affects the value update during the training, and $\lambda$ is the learning parameter.

- We assume $0 \leq \lambda_T < \lambda_C \leq 1$. i.e. the intervention gets expectations closer to the true value of the job.

- Let $Z$ is a binary treatment assignment indicator, which is orthogonal to $R$, $V_0$ and $\varepsilon$.

- Let $D$ is a binary indicator for training completion and placement.

- Let $S$ is a binary indicator for the individual staying in the DDU-GKY job for at least 5 months.

Timing:

- t=0: individual decides to go to the DDU-GKY centre iff $V_0 > R$.

- t=1: individuals decide to complete the training and take up the job ($D = 1$) iff $V_1 > R$.

- t=2: individuals who are placed learn about the true value and decide to stay in the job for at least five months ($S = 1$) iff $R < 0$.

## 3.2 Treatment effect on training completion and placement

**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 treatment discourages over-optimistic candidates and encourages over-pessimistic ones.*

**Proof**: Given $(D = 1)$ iff $V_1 > R$, the above becomes

$$P(V_1 > R|Z = 1) - P(V_1 > R|Z = 0) = P(\lambda_T V_0 + \varepsilon > R) - P(\lambda_C V_0 + \varepsilon > R)$$

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

$$P(V_1 > R|Z = 1, V_0 > 0) - P(V_1 > R|Z = 0, V_0 > 0) < 0$$
$$P(V_1 > R|Z = 1, V_0 < 0) - P(V_1 > R|Z = 0, V_0 < 0) > 0$$

$\square$

Figure 1 illustrates our theoretical predictions with a numerical simulation. 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 which the expected of the job is higher than outside options.

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 which $\lambda_C = .8$. 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 which $\lambda_T = .2$, the distribution of $V_1$ is much more concentrated around the true value of the job (Panel C). For over-optimistic trainees, the treatment reduces $V_1$, which increases their probability to end up dropping out. For over-pessimistic trainees, the treatment increase $V_1$, bringing it closer to the truth, which increases their probability to complete the training.

## 3.3 Treatment effect on staying in job conditional on placement

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

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

**Proof**:
$$P(S = 1|Z = 1, V_1 > R) - P(S = 1|Z = 0, V_1 > R)$$
$$= P(S = 1|\lambda_T V_0 + \varepsilon > R) - P(S = 1|\lambda_C V_0 + \varepsilon > R)$$

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

## 3.4 Treatment effect on being in the job five months after training

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

- *For people for which 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 people for which the job has a higher value than the outside option ($R < 0$) and that are overoptimistic ($V_0 > 0$), the treatment decreases the probability to be in the job five months after training by decreasing their probability to be placed.*

10

- *For people for which the job has a higher value than the outside option ($R < 0$) and that are overpessimistic ($V_0 < 0$), the treatment increases the probability to be in the job five months after training by increasing their probability to be placed.*

**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:

$$P(S = 1|Z = 1, R < 0) - P(S = 1|Z = 0, R < 0)$$
$$= P(D = 1|Z = 1, R < 0) - P(D = 1|Z = 0, R < 0)$$
$$= P(\lambda_T V_0 + \varepsilon > R|R < 0) - P(\lambda_C V_0 + \varepsilon > R|R < 0)$$

The sign depends on the sign of $V_0$:

- For people that are overoptimistic ($V_0 > 0$) 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 people that are overpessimistic ($V_0 < 0$) and have positive value of being in the job ($R < 0$) the treatment effect on the probability of being in the job is positive:

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

$\square$

## 3.5 Extension with job preparedness

By increasing trainees' awareness about the jobs early-on in the training, the intervention may prepare them better for the job. We model this as an increase in the true value of the job by $\tau$, leaving the outside option unchanged. This is equivalent, in our model, to a reduction of the value of the outside option $\tau$ for trainees in the treated group.

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

**Proof**: Treated trainees will therefore 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 $R < 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

## 3.6 Discussion

Notes about the set-up:

- The current model keeps the number of periods and hence the possibilities to dropout to a minimum. In reality, trainees can drop out any time during the training and the 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 the sake of simplicity, and also to preserve statistical power in our empirical analysis.

- In the current theoretical framework, training completion is the same as being placed. An alternative modelling choice would be to assume that trainees receive some information between training completion and placement. Even if the information received is similar in treatment and control groups, it may have a different effect on placement conditional on completion if we think that treated trainees are already more informed than control groups ones. We choose not to explore this aspect for two reasons: (i) the intervention does not come into this timeline; and (ii) this differential effect across groups is likely to be of second-order compared to the other effects.

- We assumed that only one job can be offered to trainees. In practice trainees can a priori be offered several jobs (this varies across training centres), which they may value more or less. The probability of being the placement job would be higher if better informed trainees were more likely to get jobs they preferred.

## 4 Results

### 4.1 Empirical framework

We estimate models using observations on individuals present at the baseline survey. A batch $b$ is in treatment group if $Z_b = 1$, in the control group if $Z_b = 0$. An individual $i$ is

assigned to a batch $b(i)$, which belongs to a randomisation stratum $s(i)$, and has a vector of characteristics $X_i$ (control variables). We consider the outcome $y_i$. Our main estimation model will be:

$$y_i = \beta Z_{b(i)} + X_i'\alpha + \delta_{s(i)} + \varepsilon_i.$$

$\beta$ is the intention-to-treat effect, the parameter of interest in our setting. We use post-double-selection lasso a la Belloni et al. (2013) to select which baseline characteristics are included as controls. We cluster standard errors by batches, 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 3 presents the effect of the treatment on our main outcomes. We first consider the probability to be in a job offered by DDU-GKY five months after training completion (column [1]). This is the unconditional probability and is based on the full sample. Hence, the dependent variable takes the value of 0 for those candidates who did not complete the training, and those who completed the training but not placed. In the control group, 35% of all candidates who started the training are in a placement job five months after the end of the training. In the treatment group this probability is 6 percentage points (ppt) (17%) higher. However, the coefficient is not statistically significant.

Column [2] provides the results for the effect on the probability of dropping out of the training programme. This probability is 13% in the control group, and not significantly different in the treatment group. The results for the probability of being placed among the candidates who completed the training are provided in column [3]. The probability is 52% and is not different in the two groups. The final column [4] presents the results for the effect of the intervention 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 65% in the control group, a 17% increase. The effect is positive and significant at the 5% level, but is not robust to multiple hypothesis testing à la Benjamini and Hochberg (1995).

In summary, these results suggest that treatment increases the probability of staying in the DDU-GKY job for at least 5 months by about 17%, regardless of whether we look at this conditionally or unconditionally. However, the power of the test is low. In contrast, we could not find significant effects on either the probability of dropout, or on the probability of placement conditional on dropout. Within the framework of our theoretical model, these

findings are consistent with the selection effects cancelling out on average. For example, the increase in dropout among candidates who are poor fits for the job and the decrease in dropout among candidates who are good fits for the job, cancelling out. They are also consistent with a positive job readiness effect to the extent that it is not anticipated, e.g. that it has no positive effect on dropout or placement but only materialises once the candidates are actually on the job.

In addition to the main outcomes, the endline surveys collect information on broader outcomes: trainees' employment situation, their location, their life satisfaction, and whether they use skills from training in their current employment if they were employed. The treatment effects on these additional outcomes are presented in Table 4. We do not find any evidence in support that the intervention increases formal employment among trainees: although the estimated treatment effect is about 10% of the control mean, it is not significant (p-value of 0.31). This suggests that some of the positive treatment effect on the probability to be in the placement job after five months, is compensated by trainees in the control finding other paid employment.

## 4.3 Heterogeneity

Tables 5 and 6 present treatment effects on the main outcomes for subsamples defined by sex (females vs. males), caste (Schedule caste/Schedule tribe vs. 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 was above or below the median of the realised placement salary).[4] These dimensions of heterogeneity were pre-specified. Caste and sex correspond to two socio-demographic dimensions of focus for the DDU-GKY policy. Education and expectations correspond to the two dimensions of heterogeneity in 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 actually does, may be more likely to be disappointed when they get the offer.

We first consider the treatment effects for females and males separately. As Table 5 Panel A shows, in the absence of any intervention, compared to males, females are much more likely to be placed in a DDU-GKY job (64% vs. 35%) and to be still working in that job five months after completing their training (47% vs. 20%). The experiment clearly shows differential treatment impact by sex. However, the intervention was not found to have any significant effect on females. In contrast, the intervention is estimated to have a large positive impact on the probability to be placed and to remain in the job, for males. The effects

---

[4]The median calculated within strata (state×trade)

are significant at conventional significance levels. They are also very large in economic terms: the intervention increases the probability of placement of men by 55%, from 20 to 31%, i.e. it closes around a third of the gender gap in placement and staying in job.

We next turn to the heterogeneity in treatment effects along the caste dimension. As Panel B in Table 5 shows that, on average, trainees from disadvantaged background (SC/ST) are more likely to be placed (60% vs. 45%) and also to be still in the DDU-GKY job (40% vs. 31%) after 5 months than trainees from OBC/General castes. The heterogeneity of the treatment effect is not very strong in this dimension but tends to exacerbate the differences across castes: the probability of staying in the placement job conditional on being placed increases by 17 ppt for SC/ST trainees (p-value of 0.012), as compared to 5 ppt for higher-caste ones (p-value of 0.39), the p-value of the difference between the two effects is 0.12.

In Panel A in Table 6, we study the treatment heterogeneity by trainee's highest educational attainment. This classification for more or less education is based on whether the individual has completed 12th grade or higher, or not. As expected, in the control, conditional on training completion, less educated trainees are more likely to be placed and stay in the job than more educated ones, which is consistent with the fact that the jobs available are only semi-skilled. However, less educated trainees drop out twice as often as more educated ones (19% vs. 9%), which may be due to difficulties learning the content of the training program. Interestingly, the intervention is found to reduce dropout for the less educated by 35% and increase the dropout for the more educated trainees by 55%. The difference between the two effects is highly significant (p-value 0.002). This suggests that better information improves the fit between candidates and jobs. The intervention, is not found to affect placement rates for any of the two groups conditional on training completion, but found to increase the probability to stay in DDU-GKY jobs for both groups, although the effect is stronger and significant only for the less educated candidates.

Finally, Panel B in Table 6 consider heterogeneity in terms of expected placement salary at baseline. Interestingly, majority of trainees (1491 of them) had expectations higher than the median wage that was offered at the end of the training: only a quarter (401 candidates) had over-pessimistic expectations, which is consistent with the idea that over-pessimistic candidates are less likely to join the programme. As expected, over-pessimistic candidates are also more likely to dropout (15%) and less likely to accept the job offer (43%) than over-optimistic candidates in the control group. There is some evidence of a decrease in dropout and an increase in placement for the low expectations candidates, which goes in the expected direction, but given the small sample size neither effect is significant.

15

We next turn to the discussion of possible mechanisms to explain these findings.

## 4.4 Mechanisms

Following our theoretical framework, our intervention could have two distinct effects on outcomes for the trainees: a job readiness and a selection effect. The selection effect comes from the fact that the intervention delivers information about jobs. Better-informed trainees make more 'accurate' 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 effect comes from the fact that the intervention prepares candidates better for the transition to employment, so that they are more likely to stay in the job once placed.

Our results are more consistent with the selection mechanism. On the one hand, the increase in training dropout among trainees who are more educated and the decrease in dropout among trainees who are less educated, are consistent with the selection story. More educated trainees would have better outside options than the jobs offered to DDU-GKY trainees, and would have not stayed in the job anyway, whereas less educated trainees may face difficulties in completing the training but are a better fit for the jobs. On the other hand, the strong positive effect on the probability to stay in the DDU-GKY job conditional on being placed without any effect on dropout or placement on average is inconsistent with greater job preparedness. These effects could still be interpreted with improved job preparedness, but only if this effect is not anticipated by the trainees, i.e. if it does not change their decision to continue training or accept placement.

In order to shed some light on the mechanisms, we study directly the effect of the intervention on the expectations of trainees regarding their labour market prospects post-training. We do this by using the information collected at the midline survey that was carried out before the end of the training and the second part of the intervention. An important caveat is that 27% of trainees were absent at the time of the midline survey including those who had dropped out already. This implies that the information collected in this survey is unlikely to capture the mechanisms highlighted in the model. In particular, if the treatment lower expectations of over-optimistic trainees and makes them drop-out or if it increases expectations of over-pessimistic trainees and make them stay, average expectations in the sample of trainees still enrolled may not change.

With this caveat in mind, Tables 7 and 8 present treatment effects on candidate's expectations. The results do not support an effect of treatment on expectations of trainees. In Table 7 we do not find any significant effect on: (i) their expected probability of getting a job; (ii) the average wage they expect from this offer; (iii) the range in which they expect

this offer to be; (iv) the location of the job they will be offered.[5] Table 8 presents some evidence that trainees in the treatment group revise downwards their willingness to accept a job outside of their state of residence, from 86% to 83% (p-value 11%) and the likelihood that they will stay 12 months outside of state from 86 to 82% (p-value 8%). 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 trainees become more realistic about their placement outcomes.

An alternative explanation for our findings which does not fit into our model is that the intervention affected the match between trainees' preferences and the jobs offered to them. For example, it could be that better awareness of the job opportunities available at the end of the training would allow trainees to communicate their preferences to the placement officers and choose their preferred options. As Banerjee and Chiplunkar (2018) show, placement officers in DDU-GKY lack information about trainees' preferences and better information given to them leads to better (more durable) matches. To investigate this possibility, we decompose the placement process in three different steps: job offer, offer acceptance and job placement as defined in our main outcomes (for a duration of two months), and estimate the treatment effect on each of them, within the sample of trainees who completed the training. Table A1 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 effect on the likelihood that trainees stay in the job two months (p-value 0.091), and that conditional on staying two months they stay at least five months (p-value 0.046). These results confirm that the treatment improves the fit between trainees and the 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 report results from an experiment that provided information on job prospects to randomly selected batches of trainees from the vocational training programme DDU-GKY in Bihar and Jharkhand (India). We find that better informed trainees are more likely to stay in the jobs they were placed in. The increase in the probability of staying in the placed

---

[5]For completeness, we report in Appendix Tables A2 and A3 results for the different dimensions of heterogeneity (gender, caste, education and expectations at baseline) but not find any significant effect by subgroup or any differential effect. In particular, there is a positive effect on salary expectations for trainees who were over-pessimistic at baseline and a negative effect on salary expectations for trainees who were over-optimistic at baseline, but none of these effects is significant.

job for at least 5 months seems to be driven by self-selection, i.e., better information on the jobs offered at the end of the training led some trainees who had better outside options, to dropout earlier, and also motivated other trainees for whom the job opportunities were attractive, to complete the training. For the average trainee, the probability of being in a job increased with no observable change in dropout or placement. These findings are inconsistent with greater motivation or greater preparedness for the transition to employment. The positive effect of the intervention on job retention are modest on average, but very large for male candidates, with a 55% increase in the probability of staying in the DDU-GKY job for at least five months on completion of the training. Given the small cost of the intervention, this suggests that detailed information about job prospects should be delivered as part of the curriculum of vocational training programmes.

# References

ABEBE, G., S. CARIA, M. FAFCHAMPS, P. FALCO, S. FRANKLIN, S. QUINN, AND F. SHILPI (2017a): *Job fairs: Matching firms and workers in a field experiment in Ethiopia*, The World Bank.

ABEBE, G., S. CARIA, AND E. ORTIZ-OSPINA (2017b): "The Selection of Talent," Working paper.

ABEBE, G. T., S. CARIA, M. FAFCHAMPS, P. FALCO, S. FRANKLIN, AND S. QUINN (2018): "Anonymity or distance? Job search and labour market exclusion in a growing African city," CEPR Discussion Paper No. DP13136.

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*.

ALZÚA, M. L., S. BATBEKH, A. BATCHULUUN, B. DALKHJAV, AND J. GALDO (2020): "Demand-Driven Youth Training Programs: Experimental Evidence from Mongolia," *The World Bank Economic Review*, lhaa013.

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 (2019): "Screening and Signaling Non-Cognitive skills: Experimental evidence from Uganda," USC-INET Research Paper 19-08.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2013): "Inference on Treatment Effects after Selection among High-Dimensional Controls†," *The Review of Economic Studies*, 81, 608–650.

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.

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

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.

ORKIN, K., E. CARRANZA, R. GARLICK, AND N. RANKIN (2020): "Job Search and Hiring with Two-sided Limited Information about Workseekers' Skills," CSAE Working Paper Series 2020-10, Centre for the Study of African Economies, University of Oxford.

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.

Table 1: Attrition rates

| Survey Timing | Control Mean | Treatment Effect | p-value | Obs. |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| Midline | 0.268 | -0.024 | 0.305 | 2488 |
| First Endline | 0.041 | -0.003 | 0.662 | 2488 |
| Second Endline | 0.049 | -0.001 | 0.886 | 2488 |
| COVID-related Sample Restriction | 0.202 | 0.007 | 0.868 | 2488 |

Notes: Column [2] is obtained from the regression of attrition dummy on an intercept and the treatment indicator, controlling for the strata fixed effects. The p-values associated with the test of no effect of treatment, are given in Column [3]. The total number of observations used is 2488.

Table 2: Baseline summary statistics (averages) and balance test - [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 Appendix Table A5. 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 1976.

Table 2: 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 | 1041 | 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 2: Baseline summary statistics (averages) and balance test (cont'd) [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×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].

Table 3: 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 | 0.060 | 0.004 | 0.019 | 0.107 |
| (standard error) | (0.044) | (0.022) | (0.050) | (0.053) |
| p-value | 0.180 | 0.848 | 0.708 | 0.047 |
| q-value (MHT) | 0.360 | 0.849 | 0.849 | 0.188 |
| | | | | |
| Observations | 1892 | 1909 | 1647 | 825 |
| Control Mean | 0.353 | 0.133 | 0.515 | 0.651 |
| Sample | All | All | Trained | Placed |

Notes: See Appendix Table A4 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 lasso selection (Belloni et al., 2013) 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 4: 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 | -0.002 | -0.023 | -0.058 | -4.209 |
| (standard error) | (0.047) | (0.060) | (0.069) | (2.610) |
| p-value | 0.965 | 0.702 | 0.406 | 0.111 |
| q-value (MHT) | 0.966 | 0.937 | 0.812 | 0.446 |
| | | | | |
| Observations | 1909 | 1908 | 888 | 1909 |
| Control Mean | 0.480 | 0.454 | 0.825 | 73 |
| **Panel B: Five Months after Training** | | | | |
| Treatment | 0.040 | 0.022 | 0.016 | 5.276 |
| (standard error) | (0.039) | (0.045) | (0.062) | (3.886) |
| p-value | 0.313 | 0.632 | 0.797 | 0.182 |
| q-value (MHT) | 0.627 | 0.797 | 0.797 | 0.627 |
| | | | | |
| Observations | 1892 | 1892 | 810 | 1044 |
| Control Mean | 0.421 | 0.367 | 0.711 | 71 |

Notes: See Appendix Table A4 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 a 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 lasso selection (Belloni et al., 2013) 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 5: Heterogeneity of treatment effects by gender and caste

| | In Placement Job after 5m (unconditional) | Training Dropout | Job Placement (conditional) | In Placement Job after 5m (conditional) |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| **Panel A: Gender** | | | | |
| Treatment * Female | 0.014 | -0.013 | -0.094 | 0.074 |
| | (0.056) | (0.025) | (0.064) | (0.049) |
| Treatment * Male | 0.109 | 0.021 | 0.129 | 0.125 |
| | (0.056) | (0.031) | (0.069) | (0.109) |
| | | | | |
| p-value Treatment Female | 0.808 | 0.606 | 0.146 | 0.141 |
| p-value Treatment Male | 0.057 | 0.508 | 0.066 | 0.257 |
| p-value Difference | 0.220 | 0.395 | 0.018 | 0.665 |
| | | | | |
| Observations | 1892 | 1909 | 1647 | 825 |
| Control Mean Female | 0.474 | 0.118 | 0.635 | 0.745 |
| Control Mean Male | 0.197 | 0.151 | 0.353 | 0.423 |
| **Panel B: Caste** | | | | |
| Treatment * Lower Caste | 0.089 | -0.003 | 0.034 | 0.172 |
| | (0.070) | (0.039) | (0.061) | (0.067) |
| Treatment * Higher Caste | 0.047 | 0.008 | 0.014 | 0.054 |
| | (0.042) | (0.029) | (0.051) | (0.062) |
| | | | | |
| p-value Treatment Lower Caste | 0.209 | 0.930 | 0.584 | 0.012 |
| p-value Treatment Higher Caste | 0.275 | 0.795 | 0.779 | 0.394 |
| p-value Difference | 0.531 | 0.827 | 0.700 | 0.117 |
| | | | | |
| Observations | 1892 | 1909 | 1647 | 825 |
| Control Mean Lower Caste | 0.405 | 0.134 | 0.595 | 0.683 |
| Control Mean Higher Caste | 0.311 | 0.132 | 0.450 | 0.617 |

Notes: The dependent variables are all binary indicators taking the value of 1 as follows. Column [1]: The trainee was in 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 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, "Higher Caste" a dummy variable for Other Backward Castes and General Castes. All regressions control for baseline characteristics chosen by lasso selection (Belloni et al., 2013) 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 6: Heterogeneity of treatment effects by education and baseline expectations

| | In Placement Job after 5m (unconditional) | Training Dropout | Job Placement (conditional) | In Placement Job after 5m (conditional) |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| **Panel C: Education** | | | | |
| Treatment * Less Educated | 0.084 | -0.063 | 0.027 | 0.141 |
| | (0.058) | (0.034) | (0.058) | (0.060) |
| Treatment * More Educated | 0.050 | 0.054 | 0.014 | 0.074 |
| | (0.045) | (0.022) | (0.055) | (0.064) |
| | | | | |
| p-value Treatment Less Educated | 0.151 | 0.064 | 0.647 | 0.022 |
| p-value Treatment More Educated | 0.266 | 0.019 | 0.805 | 0.253 |
| p-value Difference | 0.501 | 0.002 | 0.813 | 0.317 |
| | | | | |
| Observations | 1892 | 1909 | 1647 | 825 |
| Control Mean Less Educated | 0.398 | 0.186 | 0.607 | 0.680 |
| Control Mean More Educated | 0.318 | 0.090 | 0.449 | 0.622 |
| **Panel D: Salary Expectations** | | | | |
| Treatment * Low Expectations | 0.069 | -0.020 | 0.047 | 0.117 |
| | (0.092) | (0.032) | (0.099) | (0.153) |
| Treatment * High Expectations | 0.063 | 0.005 | 0.014 | 0.103 |
| | (0.044) | (0.025) | (0.052) | (0.048) |
| | | | | |
| p-value Treatment Low Expectations | 0.456 | 0.536 | 0.636 | 0.447 |
| p-value Treatment High Expectations | 0.159 | 0.838 | 0.790 | 0.037 |
| p-value Difference | 0.948 | 0.510 | 0.732 | 0.923 |
| | | | | |
| Observations | 1892 | 1909 | 1647 | 825 |
| Control Mean Low Expectations | 0.266 | 0.106 | 0.349 | 0.547 |
| Control Mean High Expectations | 0.371 | 0.138 | 0.550 | 0.665 |

 Notes: The dependent variables are all binary indicators taking the value of 1 as follows. Column [1]: The trainee was in 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 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, "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 lasso selection (Belloni et al., 2013) 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 7: Results: Intermediary Outcomes

| | Likelihood Job Offer | Average Salary | Max - Min Salary | Likelihood Job Offer Out of State |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| Treatment | 0.033 | -164.307 | -127.863 | 0.026 |
| (standard error) | (0.090) | (272.409) | (224.540) | (0.148) |
| p-value | 0.717 | 0.548 | 0.571 | 0.861 |
| q-value (MHT) | 0.862 | 0.862 | 0.862 | 0.862 |
| | | | | |
| Observations | 1486 | 1486 | 1486 | 1486 |
| Control Mean | 9.426 | 11217.0 | 3737.0 | 8.876 |

Notes: The dependent variables are captured at midline survey. See Appendix Table A4 for variable definitions. Likelihood variables range from 0% to 100%. Column [1]: Likelihood of getting a job at the end of the training; Column [2]: Expected average salary on the job offered at the end of the training (in rupees); Column [3]: The difference between the maximum and the minimum expected salary on the job offered at the end of the training (in rupees); Column [4]: Likelihood of getting a job outside of state at the end of the training. All regressions control for baseline characteristics chosen by lasso selection (Belloni et al., 2013) as well as strata fixed effects. Standard errors account for clustering at the batch level.

## Table 8: Results: Secondary Outcomes

| | Expected Earnings in 12 Months | Preferred Earnings in 12 Months | Likelihood Training Completion | Training Useful | Training Satisfaction |
|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] |
| Treatment | 428.528 | 1008.218 | 0.059 | 0.001 | 0.074 |
| (standard error) | (505.077) | (608.767) | (0.038) | (0.081) | (0.073) |
| p-value | 0.399 | 0.102 | 0.125 | 0.990 | 0.315 |
| q-value (MHT) | 0.514 | 0.282 | 0.282 | 0.991 | 0.473 |
| | | | | | |
| Observations | 1486 | 1486 | 1486 | 1486 | 1486 |
| Control Mean | 14436.375 | 18036.502 | 9.807 | 9.483 | 9.515 |

| | Likelihood Accept Job Out of State | Likelihood Stay 12 Months Out of State | Likelihood Accept Job In State | Likelihood Stay 12 Months In State | |
|---|---|---|---|---|---|
| | [6] | [7] | [8] | [9] | |
| Treatment | -0.297 | -0.351 | 0.004 | -0.180 | |
| (standard error) | (0.182) | (0.203) | (0.160) | (0.172) | |
| p-value | 0.106 | 0.088 | 0.979 | 0.300 | |
| q-value (MHT) | 0.282 | 0.282 | 0.991 | 0.473 | |
| | | | | | |
| Observations | 1486 | 1486 | 1486 | 1486 | |
| Control Mean | 8.644 | 8.601 | 8.738 | 8.643 | |

 Notes: The dependent variables are captured at midline survey. See Appendix Table A4 for variable defini-
tions. Likelihood variables as well as training usefulness and satisfaction are expressed as range from 0% to
100%. Column [1]: Expected earnings after 12 months; Column [2]: Desired earnings after 12 months; Column
[3]: Likelihood of completing the training; Column [4]: Degree to which the training is useful; Column [5]:
Degree to which the trainees are satisfied with the training; Column [6]: Likelihood of accepting a job outside
the state; Column [7]: Likelihood of staying 12 months in a job outside the state after accepting it; Column [8]:
Likelihood of accepting a job in the state; Column [9]: Likelihood of staying 12 months in a job in the state
after accepting it.
All regressions control for baseline characteristics chosen by lasso selection (Belloni et al., 2013) as well as strata
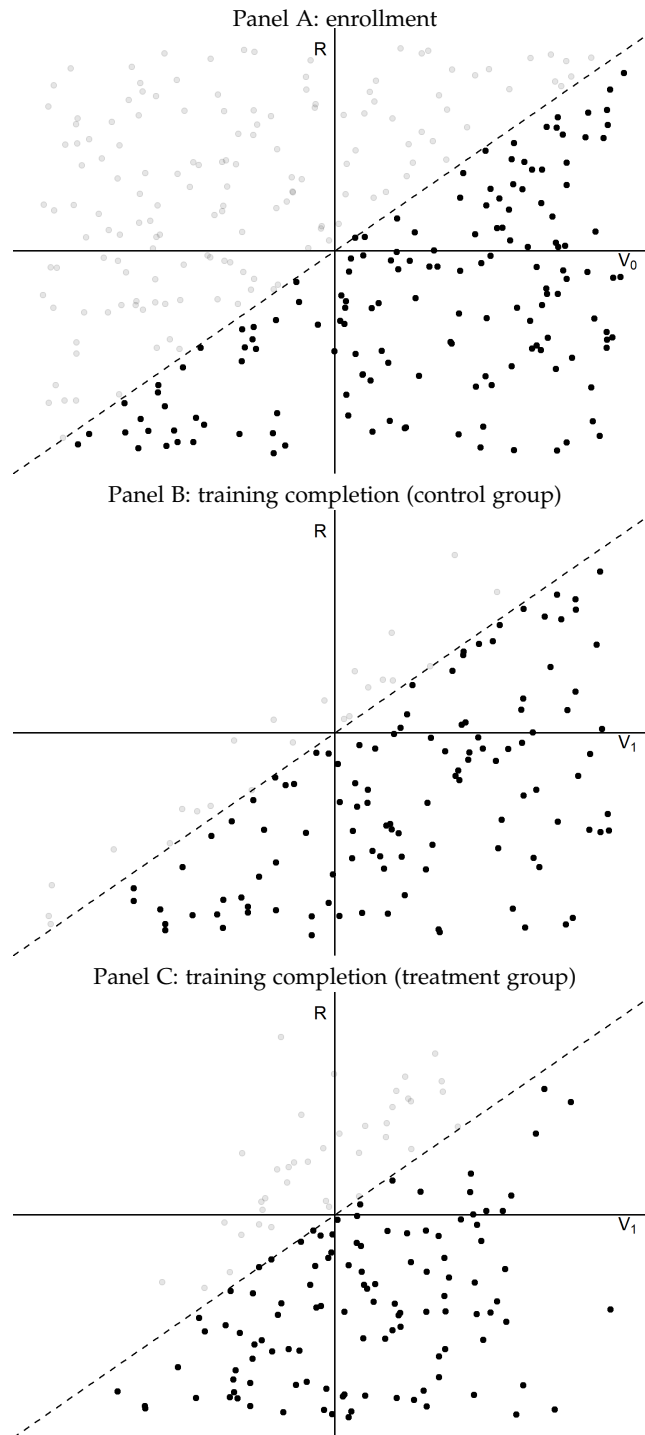fixed effects. Standard errors account for clustering at the batch level. The reported p-value is for the test of
no treatment effect and the q-value is the p-value of the same test accounting for multiple hypothesis testing
(MHT) following the False Discovery Rate method by Benjamini and Hochberg (1995).

Figure 1: Theoretical framework: numerical illustration

Panel A: enrollment



Panel B: training completion (control group)



Panel C: training completion (treatment group)

# Appendix

## A   Other outcomes

Table A1: Results: Job offer, acceptance, placement, and staying in job

|  | Job Offer | Job Acceptance | Job Placement | In placement job after 5m |
|---|---|---|---|---|
|  | [1] | [2] | [3] | [4] |
| Treatment | -0.023 | -0.032 | 0.086 | 0.108 |
| (standard error) | (-0.05) | (-0.041) | (-0.05) | (-0.053) |
| p-value | 0.646 | 0.438 | 0.091 | 0.046 |
| q-value (MHT) | 0.646 | 0.585 | 0.182 | 0.182 |
|  |  |  |  |  |
| Observations | 1536 | 1267 | 1014 | 819 |
| Control Mean | 0.844 | 0.826 | 0.792 | 0.649 |

Note: 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 job offer being made; Column[3]: The trainee was placed in a DDU-GKY job conditional on job acceptance; Column[4]: The trainee was still in DDU-GKY job conditional on placement. All regressions control for baseline characteristics chosen by lasso selection (Belloni et al., 2013) 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 A2: Heterogeneity of treatment effects on intermediary outcomes by gender and caste

| | Likelihood Job Offer | Average Salary | Max - Min Salary | Likelihood Job Offer Out of State |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| Treatment * Female | 0.024 | 92.246 | -68.566 | 0.027 |
| | (0.130) | (381.709) | (323.169) | (0.209) |
| Treatment * Male | 0.050 | -399.054 | -141.888 | -0.014 |
| | (0.112) | (323.021) | (262.927) | (0.205) |
| | | | | |
| p-value Treatment Female | 0.853 | 0.810 | 0.833 | 0.899 |
| p-value Treatment Male | 0.655 | 0.221 | 0.591 | 0.944 |
| p-value Difference | 0.878 | 0.315 | 0.860 | 0.883 |
| | | | | |
| Observations | 1486 | 1486 | 1486 | 1486 |
| Control Mean Female | 9.379 | 10735.991 | 3851.982 | 8.866 |
| Control Mean Male | 9.490 | 11869.552 | 3582.090 | 8.890 |
| **Panel B: Caste** | | | | |
| Treatment * Lower Caste | 0.044 | -199.792 | -66.331 | 0.008 |
| | (0.136) | (328.706) | (325.516) | (0.148) |
| Treatment * Higher Caste | 0.039 | -220.929 | -227.734 | 0.016 |
| | (0.102) | (295.826) | (235.740) | (0.200) |
| | | | | |
| p-value Treatment Lower Caste | 0.749 | 0.545 | 0.839 | 0.955 |
| p-value Treatment Higher Caste | 0.707 | 0.458 | 0.337 | 0.937 |
| p-value Difference | 0.974 | 0.943 | 0.639 | 0.970 |
| | | | | |
| Observations | 1486 | 1486 | 1486 | 1486 |
| Control Mean Lower Caste | 9.435 | 11218.820 | 3825.281 | 8.994 |
| Control Mean Higher Caste | 9.418 | 11216.028 | 3665.127 | 8.778 |

Notes: The dependent variables are captured at midline survey. See Appendix Table A4 for variable definitions. Likelihood variables range from 0% to 100%. Column [1]: Likelihood of getting a job at the end of the training; Column [2]: Expected average salary on the job offered at the end of the training (in rupees); Column [3]: The difference between the maximum and the minimum expected salary on the job offered at the end of the training (in rupees); Column [4]: Likelihood of getting a job outside of state at the end of the training. "Lower Caste" is a dummy variable equal to one for Scheduled Tribes and Scheduled Caste, "Higher Caste" a dummy variable for Other Backward Castes and General Castes. All regressions control for baseline characteristics chosen by lasso selection (Belloni et al., 2013) as well as strata fixed effects. Standard errors account for clustering at the batch level.

Table A3: Heterogeneity of treatment effects on intermediary outcomes by education and baseline expectations

|  | Likelihood Job Offer | Average Salary | Max - Min Salary | Likelihood Job Offer Out of State |
|---|---|---|---|---|
|  | [1] | [2] | [3] | [4] |
| **Panel A: Education** | | | | |
| Treatment * Less Educated | -0.047 | 112.018 | -14.416 | 0.012 |
|  | (0.126) | (360.848) | (336.552) | (0.157) |
| Treatment * More Educated | 0.097 | -357.863 | -187.254 | 0.005 |
|  | (0.107) | (300.388) | (208.194) | (0.203) |
|  | | | | |
| p-value Treatment Less Educated | 0.713 | 0.757 | 0.966 | 0.937 |
| p-value Treatment More Educated | 0.371 | 0.238 | 0.372 | 0.979 |
| p-value Difference | 0.332 | 0.200 | 0.573 | 0.975 |
|  | | | | |
| Observations | 1486 | 1486 | 1486 | 1486 |
| Control Mean Less Educated | 9.444 | 10728.875 | 3710.334 | 9.064 |
| Control Mean More Educated | 9.413 | 11566.609 | 3756.739 | 8.741 |
| **Panel B: Salary Expectations** | | | | |
| Treatment * Low Expectations | 0.025 | 119.078 | -185.119 | 0.210 |
|  | (0.119) | (373.083) | (393.841) | (0.344) |
| Treatment * High Expectations | 0.034 | -198.682 | -104.237 | -0.016 |
|  | (0.107) | (303.376) | (241.127) | (0.153) |
|  | | | | |
| p-value Treatment Low Expectations | 0.836 | 0.751 | 0.640 | 0.543 |
| p-value Treatment High Expectations | 0.754 | 0.515 | 0.667 | 0.915 |
| p-value Difference | 0.956 | 0.461 | 0.846 | 0.497 |
|  | | | | |
| Observations | 1486 | 1486 | 1486 | 1486 |
| Control Mean Low Expectations | 9.522 | 10610.956 | 3544.118 | 8.691 |
| Control Mean High Expectations | 9.406 | 11343.568 | 3777.642 | 8.914 |

Notes: The dependent variables are captured at midline survey. See Appendix Table A4 for variable definitions. Likelihood variables range from 0% to 100%. Column [1]: Likelihood of getting a job at the end of the training; Column [2]: Expected average salary on the job offered at the end of the training (in rupees); Column [3]: The difference between the maximum and the minimum expected salary on the job offered at the end of the training (in rupees); Column [4]: Likelihood of getting a job outside of state at the end of the training. "Less Educated" denotes trainees with less than 12th grade, "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 lasso selection (Belloni et al., 2013) as well as strata fixed effects. Standard errors account for clustering at the batch level.

# B  Variable definition

Table A4: Variable Definition: Outcomes

| 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: This Table defines the outcome variables. Likelihood variables are scored from 0% (least likley) to 100% (most likely)

## Table A5: Variable Definition: Controls

| 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 Schedule tribe |
| Caste: OBC | Caste is Other backward caste |
| 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: This Table defines the control variables captured during baseline survey.

## Table A5: Variable Definition: Controls (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 centre 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 houshold 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 A5: Variable Definition: Controls (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 |

Note: Earning variables are dummy variables equal to one if the survey response is above the median in the stratum (state×trade). Likelihood variables are expressed as a fraction between zero and one.

# C Timing of the surveys

Table A6: Endline completion periods

|  | Time of survey | Second Endline Pre Holi | Holi - 25 Mar | Post 25 Mar | Total |
|---|---|---|---|---|---|
|  | Pre Holi | 1092 | 187 | 893 | 2172 |
| First Endline | Holi - 25 Mar | 0 | 0 | 62 | 62 |
|  | Post 25 Mar | 0 | 0 | 254 | 254 |
|  | Total | 1092 | 187 | 1209 | 2488 |

Note: This table shows the time periods for the first and second endlines. Column 1 presents the survey periods, column 2-4 presents the surveys done during the respective periods.

# D Batches surveyed

Table A7: Bihar Batches surveyed

| Sector | Control | Treatment | Total |
|---|---|---|---|
| APPAREL | 3 | 3 | 6 |
| BFSI | 3 | 3 | 6 |
| CONSTRUCTION | 7 | 7 | 14 |
| HEALTHCARE | 1 | 1 | 2 |
| IT | 6 | 5 | 11 |
| LOGISTICS | 1 | 1 | 2 |
| RETAIL | 4 | 4 | 8 |
| TOURISMHOSPITALITY | 3 | 2 | 5 |
| TOTAL | 28 | 26 | 54 |

Note: This table shows the sector wise control and treatment batches surveyed in state of Bihar

Table A8: Jharkhand Batches surveyed

| Sector | Control | Treatment | Total |
|---|---|---|---|
| APPAREL | 9 | 8 | 17 |
| AUTOMOTIVE | 1 | 1 | 2 |
| CONSTRUCTION | 4 | 5 | 9 |
| HEALTHCARE | 1 | 1 | 2 |
| SECURITY | 1 | 1 | 2 |
| TOTAL | 16 | 16 | 32 |

Note: This table shows the sector wise control and treatment batches surveyed in state of Jharkhand