

The right person for the right job: Can information about job prospects reduce dropout in a large Indian training programme?

Pre-Analysis Plan

Bhaskar Chakravorty

Wiji Arulampalam

Clément Imbert

Roland Rathelot*

February 26, 2020

Abstract

DDU-GKY is a large-scale training programme in India. It is targeted at the rural poor youth, and combines short-term training with guaranteed job placement. Dropout during training and the lack of retention in placement jobs are the main problems faced by the programme. Based on evidence gathered in a previous project, we hypothesise that an important determinant of dropout is misinformation of trainees about the programme and placement jobs. We design an information intervention that takes the form of two information sessions, one before the training starts and the second before trainees are placed with employers for on-the-job training. We randomise these interventions across 90 training batches in Bihar and Jharkhand. We intend to use administrative and survey data to analyse the effects of these interventions on training completion, job uptake and retention, as well as other outcomes describing the labour-market and socio-economic situation of trainees.

*All authors: University of Warwick. Imbert and Rathelot: CEPR and J-PAL. We thank Apurav Yash Bhatiya, Gaurav Chiplunkar, Santhosh Mathew, Rohini Pande, Simon Quinn, Chris Woodruff for useful comments at many stages of the projects. This project has been funded by International Growth Centre, EQUIP-ESRC, CAGE-University of Warwick, and British Academy. We also thank Rishi Kishore, Avash Kumar, Kritarth Jha, Poonam Panda, Todor Tochev for excellent research assistance.

1 Introduction

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

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

DDU-GKY is implemented in 29 states and union territories. It currently has over 1426 projects being implemented by over 649 partners, in more than 552 trades from 52 industry sectors. Over 920,000 candidates have been trained and over 490,000 candidates have been placed in jobs as of December 30, 2019.¹

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

We conduct an experimental study using a randomised controlled trial (RCT), to evaluate two interventions aimed at increasing training completion, as well as post-training-job uptake and retention. The first intervention is administered at the beginning of the training, and the second one towards the end of the classroom training and before placement. The first intervention provides information about the types of jobs the trainees are being trained for, and the second one offers precise information about the actual placement positions.

¹<http://ddugky.gov.in/content/about-us-0> accessed on February 25, 2020.

These interventions attempt to reduce the mismatch between candidates' expectations and the employment opportunities offered by the programme.

Using combined survey and administrative data, we will measure dropout at every stage of the training, and job uptake and retention. We expect the first intervention, carried out at the very beginning of the training, to cause some early dropout if trainees update their expectations about salary and job location. The remaining trainees are expected to experience lower attrition afterwards. We expect the second intervention to increase job uptake and job retention.

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

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

2 Research questions

From qualitative interviews, we have identified high rates of dropouts at two points: (i) during the training period before completion; (ii) after job placement. The reasons for these high rates of dropouts are not well understood, but converge around the following:

- Some trainees initially lack knowledge about the content, the structure, and the timeline of the programme. In particular, they are not fully informed about the fact that the initial classroom-training leads to placement into jobs.
- Trainees may lack knowledge about the placement opportunities offered within the programme (precise location, wage, job title, etc.), and only learn about them when they are actually placed.

Our research explores the determinants of attrition during the training period, as well as post-training job uptake and retention. We hypothesise that the lack of information about placement jobs, leads to mismatch between the trainee's expectations, and what

the programme offers, both at the training and placement stages. This mismatch issue generates dropout, which is costly for the programme, as well as for firms and workers.

We aim to address the following questions:

1. Is attrition during the training, low job uptake and retention, due to a lack of information about the programme and the characteristics of the placement job (e.g., wage, occupation, location)? Can targeted information delivered during the training phase, alleviate the lack of information, and better align their expectations, in order to lead to better training and placement outcomes?
2. Are high dropout and low job retention linked to lack of job readiness? Can an intervention delivering simple messages about the training-to-job transition increase job uptake and retention?

3 Experimental design

3.1 Interventions

There are two interventions: one at the start of the training period (A); and another just before placement (B). Their objective is to reduce dropout during the training and increase job uptake and retention. Interventions are implemented at the level of the training batch, a group of students enrolling, having classes, and graduating together.

Intervention (A).

- **Intervention:** The intervention takes place at the very beginning of the training, before “batch freezing”. Before batch freezing, trainee dropout has low cost for training providers: they have time to attract more students to meet their quantitative objectives. We gather trainees in a classroom and provide them with a list of detailed characteristics of potentially available placement jobs. The list includes job title, company name, location (city and state), and compensation package (net monthly wage and in-kind benefits). Trainees will have an opportunity to ask the placement officer, questions regarding the contents of the list.
- **Theory of change:** We expect that: (i) dropout may increase immediately after the intervention, as learning about the programme leads some trainees to update their expectations, (ii) retention in the training may increase afterwards, as the remaining trainees would start the programme with more accurate expectations.

Intervention (B).

- **Intervention:** This intervention takes place a few days before placement. We gather trainees in a classroom and provide them with a list of detailed characteristics of positions that are actually available to them for placement. The list includes job title, company name, location (city and state), and compensation package (net monthly wage and in-kind benefits). Trainees will have an opportunity to ask the placement officer questions regarding the contents of the list.
- **Theory of change:** We expect (i) that following the intervention some trainees may decide to refuse placement, as learning about the actual opportunities may lead them to update their expectations. (ii) that the remaining trainees will be better prepared to deal with the transition to placement jobs. We expect an increase in job retention.

3.2 Design

The experiment will be conducted as a cluster randomised controlled trial with two arms, consisting of: A and B together in treatment batches, and control group (no intervention). The randomisation is carried out at the batch level, stratified state wise by sector/trade. The randomisation sample consists of 90 batches in Bihar and Jharkhand. With approximately 30 candidates per batch, we expect a sample size of 2,700 candidates.

3.3 Power calculations

We consider a power of 80% and a significance level of 5%. For the main outcome (retention in employment 5 months after placement), we assume a base mean of 70% and intra-cluster correlation of 0.1. The minimum detectable effect is 10.6 percentage points with 45 clusters per group and 30 observations per cluster.

4 Data collection

4.1 Collection process

Our research relies on:

- primary data collected from four rounds of surveys of trainees from 90 training batches across both Bihar and Jharkhand and,

- data from the management information system (MIS) data from Bihar Rural Livelihood Promotion Society (BRLPS) and Jharkhand State Livelihood Promotion Society (JSLPS).

Surveys. All surveys are administered on tablets using questionnaires designed on Survey CTO platform. The baseline and midline surveys are administered in face-to-face interviews. Both endlines are administered using phone interviews. There are four rounds of surveys: baseline, midline, and two endline surveys.

- **Baseline survey:** This survey is being administered to all participants at the training inception, between first day of training and day-10 (i.e before batch freezing which usually happens on the 10th day from the batch start date). Data collectors administer the baseline questionnaire in a face-to-face interview sessions with individual trainees. The baseline questionnaire is custom designed to capture a wide range of socio-economic 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.
- **Midline Survey:** This survey interview is conducted with the trainees at the end of the classroom training and before placement. All follow-up survey rounds, interview only 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 any other reason), are not surveyed in the follow-up rounds. The midline survey mainly captures the change in expectation of the trainees and classroom dropout.
- **First and second endline survey:** These surveys are administered over the phone. The first endline is conducted approximately two months after training completion, and the second one about five months after training completion. The objective is to collect information about the post-training outcomes focussing on information related to employment and location. These are detailed below.

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

4.2 Outcomes

Main outcomes. The main outcomes are measured using endline and midline survey data.

The first outcome will be:

- Job retention five months after the end of the training (defined for all, including those that did not actually complete the training).

We will then investigate the timing of the treatment effects through the following outcomes.

- Dropout before batch freezing.
- Dropout before placement conditional on being in the frozen batch.
- Placement conditional on having completed classroom training.
- Job retention five months after training completion conditional on placement.

As a robustness check, we will report results for the same outcomes based on administrative data.

Intermediary outcomes. The midline survey elicits information about various expectations.

- Likelihood of job offer after training completion.
- Average expected salary of the job offered after training.
- Difference between maximum and minimum expected salary of the job offered after training.
- Likelihood that the job offered after training is outside the state.

Long-term outcomes. Long-term outcomes are measured in the two endline surveys.

- Formal wage employment five months after training completion.
- Location (in or outside of home state) five months after training completion.
- Training skills useful in current employment.
- Life satisfaction.

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

- Expected earnings, occupation, location in 12 months after completing the training.
- Desired earnings, occupation, location in 12 months after completing the training.
- Likelihood of training completion.
- Likelihood of job uptake.
- Likelihood of job retention (after 12 months).
- Expected earnings in 12 months after completing the training.
- Training usefulness.
- Satisfaction with the training.
- Reasons for dropout from training.
- Reasons for refusing placement.
- Reasons for dropping out of the job.
- Number of friends in the current place of residence and work (all friends, friends from the training, friends from the village).

4.3 Control variables

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

Socio-demographics.

- Sex.
- Age.
- Relationship with the head of the household.
- Education (grade, exam score, previous vocational training).
- Marital Status.

- Occupation, location and earnings before Training.
- Caste.
- Religion.
- Household Type (nuclear, joint).
- Number of earning members of household.
- Household earning.
- Number of household members.
- Landowning household.
- House type (kuccha, pukka).
- Homeownership.
- BPL card / RSBY card / SHG Membership / MGNREGA participation in household.
- Internet use.
- Past migration (own, family members).

Non-cognitive traits.

- Grit.
- Self-Esteem.
- Big 5 (Agreeableness, Conscientiousness, Extraversion, Neuroticism, Openness).
- Risk Aversion.
- Life goals (family, work).
- Time to complete the survey.

Baseline expectations.

- Awareness about the training.
- Hypothetical earnings, occupation, location at the time of the survey if they had not joined the training.
- Hypothetical earnings, occupation, location in 12 months from the survey date if they had not joined the training.
- Expected earnings, occupation, location in 12 months after completing the training.
- Desired earnings, occupation, location in 12 months after completing the training.
- Likelihood of training completion.
- Likelihood of job offer after training completion.
- Minimum / Maximum / Average salary of the job offer.
- Likelihood that the job offer is outside the state.
- Likelihood of accepting the job in the state.
- Likelihood of staying 12 months in the job if it is in the state.
- Likelihood of accepting the job outside of the state.
- Likelihood of staying 12 months in the job outside of the state.
- Usefulness of training.
- Satisfaction with the training.
- Difficulty for the family if stay outside of the state for 12 months.

Training

- Month when training started
- Month when training ended
- District where training center is located
- Captive placement (training provider is one of the employers)
- Single placement (training provider tied with only one employer)

5 Econometric analysis

We estimate models using observations on individuals present at the baseline survey. Randomisation strata: state (Jharkhand or Bihar) times sector: total of 13 strata. A batch b is in treatment group if $T_b = 1$, in the control group if $T_b = 0$. 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).

5.1 Balance

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

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

where δ_s denote as fixed effects for strata.

We will then test the null of no difference between the treatment and control groups ($\beta = 0$). For inference, we will use randomisation inference to handle clustering and finite-sample inference issues. We will correct for multiple hypothesis testing, controlling for false discovery rates.

5.2 Main effects

We consider the outcome y_i . Our main estimation model will be:

$$y_i = \beta T_{b(i)} + X'_i \alpha + \delta_{s(i)} + \varepsilon_i.$$

β is the intention-to-treat estimate, the quantity of interest in our setting. We will use post-double-selection lasso for variable selection. For inference, we will use randomisation inference to handle clustering and finite-sample inference issues. We will correct for multiple hypothesis testing, controlling for false discovery rates.

5.3 Attrition

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

5.4 Heterogeneity

We will consider the following dimensions of heterogeneity:

- sex (female vs not).
- social background (Scheduled Caste / Scheduled Tribes vs not).
- education (completed higher secondary vs not).
- expectations (expected earnings 12 months after training).

There are several reasons why people of different sex, caste, and education would have different treatment effects:

- Women, SC-ST, and individuals with low levels of education, have lower options outside of the programme. We expect a positive effect of interventions A and B, on training completion and job uptake to be stronger for them.
- Students with higher expectations are more likely to be disappointed when the characteristics of the job are revealed to them. We expect interventions A and B to increase dropout (reduce training completion and placement) for them.

In addition to these dimensions, we will also use causal forest method for estimating heterogeneous treatment effects. This method allows one to pick the characteristics that are most relevant for explaining the heterogeneity in treatment effects.

5.5 Mechanisms

The interventions have two effects on candidates' outcomes: a direct effect and a selection effect.

- The direct effect comes from the fact that candidates who would pursue the training to the end with or without intervention are better prepared for the transition to employment and have higher job retention.
- The selection effect comes from the fact that giving information about jobs changes the decision to continue training for candidates who have different potential training and labor market outcomes.

Also, due to funding constraints, our design pools together the two interventions: intervention *A* (before batch freezing) and *B* (before placement). In general, this means that we will only be able to measure the combined impact of both interventions.

We will develop a theoretical model to clarify how to use treatment effects on outcomes at different points in the training process to test whether these different effects exist.

- A significant treatment effect on dropout before batch freezing will suggest that intervention *A* has a selection effect.
- A significant treatment effect on placement conditional on completing classroom training will suggest that intervention *B* has a selection effect.

References

- BLATTMAN, C. AND L. RALSTON (2017): "Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs," SSRN 2622220.
- BLOOM, N., G. FISCHER, I. RASUL, A. RODRIGUEZ-CLARE, T. SURI, C. UDRY, E. VERHOOGEN, C. WOODRUFF, AND G. ZANE (2014): "Firm Capabilities and Economic Growth," IGC Evidence Paper.
- 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.