How important are matching frictions in the labor market? Experimental & non-experimental evidence from a large Indian firm*

Abhijit V. Banerjee† Gaurav Chiplunkar‡

November 2020

Abstract

This paper provides evidence of substantial matching frictions in the Indian labor market. In particular, placement officers in vocational training institutes have very little information about the job preferences of candidates who they are trying to place in jobs. We begin by adopting several methods to elicit genuine preferences of candidates over different types of jobs and show that: (a) there is a substantial variation in preferences over the same jobs and (b) placement officers have poor knowledge of it. We then provide placement officers with this information and examine its impact on placement and employment outcomes. We find that placement officers come close to efficiently matching candidates to job interviews and that there is an overall improvement in matching even after taking into account redistribution within a group of potential employees. Lastly, we find substantial improvement in the quality of jobs that the treated candidates end up with, as measured by their preferences.

JEL codes: J22, J24, J61, J62, J64

Key words: Matching, labor, job-search, recruitment, experiment

*We thank Pramod Bhasin, Bruno Crepon, Esther Duflo, Glenn Ellison, Clement Imbert, Parag Pathak, Thomas Piketty, Debraj Ray and seminar participants at the Paris School of Economics, Indian School of Business (Hyderabad), Indian Institute of Management (Bangalore) for helpful comments and conversations. We thank Chetana Sabnis, Shivani Agrawal and Akankshita Dey for providing excellent research assistance and thank Skills Academy for their co-operation in implementing the field activities. We are indebted to Putul Gupta for her exceptional work in leading and managing the field implementation team. We gratefully acknowledge research grant support from PEDL Exploratory Grants (#3545) and MIT. This RCT was registered in the American Economic Association Registry for randomized controlled trials under trial number AEARCTR-0001835.

†Ford Foundation International Professor of Economics, MIT. Room 540, Morris and Sophie Chang Building, 50, Memorial Drive, Cambridge, MA 02142. Email: banerjee@mit.edu.

‡Assistant Professor of Business Administration, Darden Business School, University of Virginia. 100, Darden Blvd. Charlottesville, VA 22905. Email: chiplunkarg@darden.virginia.edu.

1 Introduction

Poor matching between job-seekers and the jobs they are allocated by those responsible for job placement is one potential explanation for high youth unemployment rates in a developing economy like India, where there seems to be no dearth of low-skilled jobs. This paper provides evidence from a field experiment that those responsible for placement lack information about the preferences of the job-seekers. As a result, we show that the matching to jobs is quite inefficient and providing the placement officers with preference information improves matching substantially and significantly.

The point of departure of this study is an important but under-emphasized fact about the Indian economy captured in figure 1(a). It uses data from the 68th round of the National Sample Survey (2011-12), which is a nationally representative survey in India and reports the non-employment rates for men at different ages for those with ten or more years of competed education and those with eight or less. The figure shows a remarkable divergence between the more and less educated categories: at age 25 for example, 31.7% of the more educated young men are not employed while the same number among the less educated men is 4%. This difference of 27.7 pp is large and statistically significant at the 0.001 level. Figure 1(b) then shows the non-employment rate between education and seeking employment. About half of the more educated males who are not working at age 25 (42.9% to be exact, which is 13.6% of the population of that cohort), claim to be available for work (though perhaps not all of them are actively looking for a job). Among the rest of the non-working educated males, almost all of whom claim to be studying, a significant fraction are actually preparing to take gateway exams that would qualify them for specific jobs (in the government, in the banking sector, etc.).[1] Taken together this is a very large population of job-seekers. Interestingly, there does not seem to be a dearth of jobs per se. At age 40, there is essentially no statistical difference between the non-employment rates of the education groups– both have non-employment rates of around 0.2% (p=0.62).[2]

The above facts suggest two possible hypotheses for why there are so many educated job-seekers. First, the search mechanism could be inefficient as it takes a long time for these job-seekers to find the job they want. Second, they start by aiming high in the job market and slowly adjust their expectations based on their experience. This could be entirely rational if, for example, some jobs have lots of rents and job-seekers focus on getting one of those jobs rather than settling for a "bad" job immediately after school or college.

The idea that there may be inefficiencies in job search is well-known. Thick market externalities (Diamond (1982), Mortensen and Pissarides (1994), Acemoglu (1996, 1997)) or distortions in

[1] The normal age for graduation from college in India is 21 or 22 and that of finishing a masters degree is 24. At 25, a lot of them have finished their general education and are probably studying for the many exams that are the gateway for specific jobs.

[2] While these could be cohort specific differences, we see very similar patterns in figure A1 in the appendix, which reports on the two previous rounds of the survey that collected the same data (the 66rd and 64th rounds from 2009-10 and 2007-08 respectively).

the job search process make it possible that a job seeker searches too little, which would justify incentivizing job search. On the other hand, job-seekers may not know how and where to search and therefore, it may be useful to provide them with external job search assistance. Both these strategies, incentives for job search and job search assistance, are reasonably common practice in OECD countries. Card et al. (2010) in their meta-analysis of active labor market policies, report on 857 separate impact estimates of which 15% come from interventions that target search behavior either through incentives or through search assistance. These are almost entirely from the OECD. More recently, interventions to reduce search and information frictions between workers and firms in developing countries have been studied by Dammert et al. (2015), who provide information on vacancies to job-seekers; Beam (2016); Abebe et al. (2017), who test the impact of job fairs; Banerjee and Sequeira (2020); Franklin et al. (2015), who subsidize job search; Abel et al. (2016); Groh et al. (2015); Bassi et al. (2017); Pallais (2014), who reduce screening costs through reference letters, skill report cards and referrals.

This paper reports on a randomized trial of an intervention that addresses a friction in the job search process that has not been much studied. We start by providing detailed evidence for an important source of mismatch in the job placement process of a large Indian vocational training firm. Placement managers (who are responsible for matching job-seekers to interviews) we show, often have little information about the job preferences of the candidates that they are responsible for placing, and as a result often offer candidates interviews for jobs that these candidates have no interest in. To document this mismatch, we need to reliably know the preferences of each job-seeker. Otherwise, what we may believe to be a mismatch, could in fact reflect that the placement manager knows more about job-seeker preferences than we do. Unfortunately, getting people to reliably reveal their preferences is not easy, especially when preferences are multi-dimensional so that standard BDM mechanism cannot be used. To elicit preferences of job-seekers over a set of job characteristics, potential job-seekers (who are currently trainees at a vocational training center) are asked to make choices by ranking a list of real-world job options. The jobs that they rank are carefully chosen to resemble the real-world jobs that they could potentially get, as well as exploit variation along different job characteristics. To test whether these rankings reflect their true underlying preferences, for half the job-seekers, we emphasize that the probability they will get an interview for a highly ranked job is high (given our partnership with the training institute, it is in fact true) and for others we make it clear that the probability is quite low (which is also true). The two preference distributions we get are essentially identical, giving us some confidence that (a) we don't need strong incentives to elicit true preferences and (b) these are their actual preferences rather than what they would report strategically to maximize their chance of getting a job. Finally, we make them list the attributes of a job that they like. It turns out that the preferences revealed by just asking them this are very consistent with preferences elicited through the more elaborate job ranking exercise described previously.

Having thus confirmed that we know what the true preferences of these job-seekers are, we ask the placement manager of the training center, to predict the preferences of each trainee over the *same* set of jobs used for eliciting trainee preferences described above. Specifically, we ask the manager to pick the three best jobs (in order of preference) from that trainee's point of view. Through various measures on how the manager's ranking correlates with the trainee's ranking, we can examine the extent to which the person in charge of placement knows the preferences of the person they are placing in a job. The results are consistent with managers having lots of information about some trainees, but very little information about others. For example in section 4, we show that the manager's ordering of the three jobs perfectly correlates with the trainee's ordering in 21% cases but is the exact opposite of the trainee's ordering in 16% of cases. On average, the job picked by the manager as the best job for a particular trainee is ranked at 7.2 by the candidate himself on a scale of 1 to 11 (1 is the worst and 11 is the best). If the manager had picked at random instead, the average rank would have been 5.5 and if the manager knew the preference perfectly, the rank should have been 11. So it seems that managers do slightly better than a completely random choice, but they are far from knowing their trainee's true preferences.

Having documented the lack of knowledge of trainee preferences by their managers, the second part of the study (section 5 onward), discusses the implementation and impact of a randomized control trial. We experimentally vary the information that placement managers have about trainee preferences, where for half of the trainees in a batch (henceforth the treatment group), we provide the manager with the job rankings (preferences) for their four most preferred jobs, and not for the others (control group). We show that this intervention substantially improves the allocation of interviews as trainees in the treatment group are on average, 10.7 pp (or 48.5%) more likely to get an interview for one of their top four jobs.

The next step is to then evaluate the overall impact of the intervention. Though trainees in the treatment group benefit from the information being provided to the manager, it does not necessarily mean that the overall matching has become more efficient, since there could be displacement effects (as in Crépon et al. (2013)). To get at this, in an ideal experiment, one would randomize information on trainee preferences at the batch level instead of across trainees within a batch (as we do). Our choice of the latter, as discussed in section 6, was primarily driven by the limited resources at our disposal as well as the logistical challenges (uncertainty of batches, challenges of surveying in remote areas, etc.) faced by operating in the rural areas of Uttar Pradesh. Therefore, the limitation of our design (that we could overcome by randomizing at the batch level) is that we do not have a counterfactual allocation of interviews within the batch in the absence of our intervention.

We adopt a more theory based approach described in section 6 to overcome this limitation. To generate a counterfactual allocation, we first try to model the decision-making rule of the placement manager. We make three possible alternative assumptions about what the manager knows–(i) a *complete information* case, where she knows what we know about the preferences of all trainees, (ii) a *no information* case, where she knows what she tells us in our manager interviews about trainee preferences, (iii) a *hybrid information* case, where she knows what we tell her for the treatment group but what she tells us for the control group. Under these alternative assumptions, we ask whether a stable matching algorithm can predict the allocation we see in the data. We find (not surprisingly) that the complete information case does a poor job at explaining the allocation of interviews and the hybrid information case fits the data the best. In other words, the manager does come close to achieving efficiency, subject to her information constraints (though why she does not ask the trainees remains an open question).[3] We then impose the assumption that in the absence of the information the manager would have allocated the interviews based on what she told us about the preferences of the job-seekers, while when she has the information we gave her, she makes use of it in addition to her prior information. Under the assumption that the allocation the manager generates is a stable match, this gives us the predicted allocations in each batch with and without the additional information. Comparing them, we see that the treatment group are 9.5 pp (or approximately 40%) more likely to get an interview for at least one of their four most preferred jobs, while those in the control group within the batch remain unaffected on average. At least by this metric, the intervention was a success.

The final section of the paper asks whether the success in altering the allocation of interviews has differential labor market consequences for the trainees. In particular, do they actually get jobs that they like better, and does that experience make them more likely to stay employed, either in the same job or in another (potentially better) job? The answer seems to be somewhat mixed. The intervention does have large and significant effects on the quality dimension of employment in the short run (three months): treated trainees get offers for jobs that they like better, accept these jobs and three months later, are more likely to be employed in them (though not after six months). On the other hand, while the impact of our intervention on the likelihood of being employed in any job (irrespective of quality) is also large and positive (16% higher after 3 months and 42% higher after 6 months), it is imprecisely estimated to be statistically significant at conventional levels. Though in Indian context, these findings are very consistent with other research studies that examine the impact of interventions matching firms and workers in other contexts like Ethiopia (Abebe et al. (2017); Blattman et al. (2019)) and the Philippines (Beam (2016)).

The rest of the paper is organized as follows. Section 2 gives some background information about the particular labor market we are studying. Section 3 then describes the methodology used to elicit preferences and what we find. Section 4 describes the results about the gap between what the trainees want and what the managers think they want. Section 5 describes the intervention, the randomized controlled trial based on it and the results. Section 6 discusses the (model-based) estimates of the general equilibrium consequences of our intervention. Section 7 reports on the impact of the treatment on various labor market outcomes and we conclude the paper in section 8.

[3] Qualitative surveys with managers after the intervention suggest that they undervalue the importance of incorporating variation in candidate preferences while allocating jobs, especially on the non-pecuniary dimensions.

2 Context and Data

2.1 Institutional setting

As discussed previously, India has a high and rising non-employment rate among the educated youth (18-29 years). At the same time, a widely cited survey on 'labor/skill shortage for industry' conducted by the Federation of Indian Chambers of Commerce and Industry (FICCI)[4] reports that 90% of firms indicate facing shortage of labor and 89% of firms report not being able to meet their potential demand in the market due to labor shortage, thus indicating (among other things) potentially a mismatch between labor demand and supply. It is therefore not surprising that active labor market policies have been at the center of policy agenda in India in the last decade.

The Government of India (as a part of the 11th Five Year Plan) launched a Skill Development Mission that initiated skill training programs under a 'Coordinated Action on Skill Development'. It proposed to integrate training efforts by various public and private entities across various sectors of the economy. The institutional structure consisted of the (i) Prime Minister's National Council on Skill Development; (ii) National Skill Development Coordination Board and (iii) National Skill Development Corporation. An ambitious targeting of training over 500 million people by 2022 was set through public-private partnerships that would be managed by the NSDC. While the NSDC designed the components of various training programs under the Skill India Mission, the private sector was incentivised to undertake their implementation through financial payouts to private training institutes after the successful completion of the training program. A crucial aspect of this financial compensation was the importance of post training placement of trainees. For the shorter 3 month training courses, 15-20% of the financial compensation was contingent on trainees being employed for three months after the competition of the training program.

On the impact of training programs in India, a study conducted by the International Labour Organization (2003) that focused on three states of Andhra Pradesh, Maharashtra and Odisha found poor labor market outcomes for the trainees after the training program. Another subsequent study by the World Bank (2008) found that a high proportion of trainees remain unemployed after the training program. Furthermore, more recent reports from the impact of training programs (NSDC (2013), FICCI (2013)) suggest two major challenges faced by trainers: first, a low take up rate of training programs and second, the tendency of trainees to quit their jobs within a short period (two-three months) of their initial job placement. Both challenges suggest a mismatch between the jobs skilling programs delivery and what their clients want. This could be either because there are not enough of the kinds of jobs the clients want or because the existing pool of jobs are not allocated to the right set of applicants.

For this study, we partner with Skills Academy[5], a large training institute that undertakes the

[4] FICCI Survey on Labor/Skill Shortage for Industry, October 2011.
[5] http://theskillsacademy.in

design, management and implementation of training programs across 17 states in India. Skills Academy focuses on training potential job-seekers in medium-level skills primarily in the service sector (hospitality, retail etc.) and placing them in jobs after the completion of the training program. A crucial aspect of the training program, which will be important for this paper is that job placements and matching to job interviews is undertaken primarily by the training center managers and as discussed above, the training institute cares about the successful placement of the trainees since a sizable fraction of the financial compensation is contingent on successful post-training placements and subsequent retention in employment.

2.2 Sample description

Our study sample consists of 538 individuals who enrolled in training programs implemented by Skills Academy across 10 centers in the states of Uttar Pradesh and the National Captial Region of Delhi. 91.26% of the sample is enrolled in three widely conducted training programs designed under the NSDC namely: the Uttar Pradesh Skill Development Mission (UPSDM), the Pradhan Mantri Kaushal Vikas Yojana (PMKVY) and Plan India. 83.7% of the trainees in our sample are enrolled in training programs that focus on healthcare, hospitality and retail sectors, while the rest are enrolled in training programs focusing on computer and automobile training. Table 1 provides the demographic description of our sample. In columns (2) and (3), we also compare our study sample to a nationally representative sample of the 68th Round of the National Sample Survey (NSS), which was conducted in 2011-12[6]. As can be seen in column (1), our study sample is young (21 years old on average), have completed their high school education and come from backward caste backgrounds. 48% of the sample is female.

3 Eliciting preferences over jobs

We now turn to eliciting preferences of trainees over job characteristics. To do this, we carried out two different exercises to learn about the job preferences of workers. We describe both of them below and then put them together to check if the two procedures give similar results.

3.1 Hypothetical choices

Job aspirations

In a survey implemented during the first week of the training program, trainees were asked about their aspirations with regard to employment after the training program. We focused specifically

[6] Skills Academy (and all government training programs) require potential trainees to be between the ages of 18 and 35, with at least a high school level of education. We therefore constrain the NSS sample to match this eligibility criteria.

on four aspects of a job that from other accounts, were important for trainees: employment sector, location, salary and whether there was provident fund (PF).[7] With regard to the sector of employment, trainees were provided with a list of seven sectors (banking, business process outsourcing or BPO, retail, hospitality, healthcare, information technology or IT and others). Trainees were then asked to rank these sectors in where they *aspire* to work in after the training program. We then create a dummy variable, which takes the value 1 for the sector that the individual most aspires to work in and report the results in panel A of table 2. 72% of the trainees report aspirations to work in the healthcare, banking and retail sectors. Next, keeping in mind their qualifications and skills, trainees were asked to describe the characteristics (salary, location and provident fund) of their *ideal* private sector job. The results for salary and provident fund are reported in panel B of table 2. Trainees report a desired salary of Rs. 15,036 on average[8], with 98% of individuals reporting a preference for a job with provident fund. Panel C reports the location preferences, which is broken down based on the residence of a trainee. For trainees in Uttar Pradesh, only 18% aspire to get a job in the local area while 74% aspire to get a job in the state capital of Lucknow or another major city in Uttar Pradesh. Only 8% are willing to move outside of the state (mainly to Delhi or Mumbai, both large metropolitan cities). For the trainees in Delhi, 97% of them want a job in Delhi while the rest are willing to move to another city.

Job priorities

In the same survey as above, trainees were asked directly about their preferences over six different job characteristics[9] by asking them to distribute a hundred points across these job characteristics. Table 3 reports the results for this activity. Column (2) reports the average points allocated by trainees to a job characteristic, while columns (4) and (5) report the values separately for males and females respectively. Lastly, column (6) reports the p-value that tests the statistical difference between columns (4) and (5). As can be seen from the table, salary, location and job title/designation are the three most important characteristics for trainees in a job and are 1.5 to 2 times more important in magnitude than other job characteristics like job security, social status and nature of work. The only significant difference across men and women is with respect to location, which perhaps not surprisingly in the Indian context, is more important for women than for men.

[7]Provident Fund is a mandatory savings scheme where a firm is required to match the employees contribution. Since only relatively established firms offer these despite the fact that all firms beyond a certain size are required to do so, offering PF might be seen as an indicator for a "good" firm.

[8]There is variation in the expected salary across states with an average of Rs. 24,373 in Delhi and Rs. 12,978 in Uttar Pradesh. When we compare this to the salary actually got through placement, the average salary in Delhi after placement is Rs. 8,176 and in Uttar Pradesh is Rs. 6,622. This difference is statistically significant at the 0.01 level.

[9]In a pilot survey, trainees reported these characteristics to be important while considering a job.

3.2 Real choices

The survey described in the previous section reports on choices made by trainees over hypothetical job scenarios. In this section, we describe an activity that presented trainees with real-world job scenarios and discusses what we learn about trainee preferences from their observed choices.

Incentivized elicitation of preferences

To begin, we first generated a list of sector-specific jobs by varying the job characteristics that trainees reported as important in the hypothetical activity above: salary, location, designation and social security. The idea of this exercise was to vary job characteristics to generate jobs that closely resembled the jobs that would be available to trainees after the completion of their training program. Salary was varied between low, medium and high categories. Provident fund was either offered or not. The job designation was varied between desk/phone jobs and activity intensive jobs. Finally, the location was varied in three ways, namely: (i) local place of residence of the trainee; (ii) large cities within the state and (iii) metropolitan cities outside the state.[10] Taking all combinations across the four characteristics would produce 36 jobs. However, we wanted to ensure that the presented jobs were as close as possible to the real world jobs. Hence, within every employment sector that a trainee was trained in, and after looking at previous jobs offered in these sectors in the past, the list of 36 jobs was narrowed down to the 11 most realistic jobs.[11] To further enhance the authenticity of the job choice exercise, it was timed to coincide with the actual placement period in the training program, which was usually in the last week of training.

At the beginning of the placement period, trainees were presented with the list of 11 jobs generated as described above, and were asked to rank them from 1 to 11 based on their preference of working in these jobs if they were offered one (1– least favorite job and 11– most favorite job). In carrying out this exercise we faced a dilemma: on the one hand, we wanted them to take the exercise seriously, which points towards making it high stakes. On the other, we wanted them to reveal their genuine preferences rather than choosing strategically to maximize their chance of getting a job. This suggested making the stakes less salient. In the end, we decided to go for the two extremes with the view that if they yielded more or less the same result, we could be reasonably confident that we have captured genuine preferences.

Specifically, within every training batch, half of the trainees, chosen at random, were told that the job ranking activity was for research purposes, and there was a very low likelihood that the job ranking exercise would influence the interviews they would get. The other half were told that there was a very high likelihood that their job rankings would determine the interviews they would get. In both cases, because of our partnership with Skills Academy, the description was factually

[10] The variation in job characteristics is summarized in table A1. For example, for the trainees in Raibareli (a town in Uttar Pradesh), location was varied between jobs in Raibareli, jobs in Lucknow (the state capital of Uttar Pradesh) and jobs in Delhi/Mumbai.

[11] See figure A2 for an example of one such list and an example of one such job.

correct. One challenge we faced in implementing this exercise however, was that since it was conducted in the last week of the training program (just prior to placements), there was irregular attendance in the training program. Therefore, despite multiple visits to the training center, we were only able to conduct the exercise for 338 trainees (63% of the sample). Table A2 shows no systematic difference in the profile of trainees who were absent on the days that this activity was conducted. For the sample of trainees for whom we have the rankings, table A3 reports a standard balance check on the observable characteristics of trainees assigned to low and high salience groups.

We now come to the results of this activity. First, in columns (2)-(4) of table 4, we report a substantial heterogeneity in the preferences that trainees have over the *same* set of jobs. For each of the 11 jobs, we calculate the fraction of trainees who placed a job in the bottom three (column (2)), in the middle i.e. between 4-8 (in column (3)) or in the top three (column (4)). For example, around a third of the trainees put jobs 2, 3, 4, 8 and 9 among their bottom three jobs, but around 20% put them in the top three. The reverse is true for jobs 6, 10 and 11. In other words, not everyone wants the same jobs. This is why there are potentially large welfare gains from reallocating jobs based on preferences. Finally, in columns (5)-(7) of table 4, we see no difference in the rank given to a job based on if a trainee was allocated to the low or high likelihood group (as described above)–the differences are both small in magnitude and nowhere near statistically significant. Going forward, we will therefore assume that these rankings reflect the true underlying preferences of trainees over jobs.

Compensating differentials

Using the reported job rankings, we can then ask how much salary are trainees willing to give up to compensate for a change in the job characteristic (keeping all other job characteristics the same). For example, we can ask, keeping all else equal, how much additional salary would a trainee desire if she were offered a job in Lucknow instead of the town of the trainee's residence. To do this, we run the following regression:

$$R_{ij} = \alpha_i + \sum_k \beta_k X_j^k + \gamma S_j + \varepsilon_{ij} \qquad (1)$$

where R_{ij} is the rank given by a trainee i to job j, X_j^k are the dummy variable for the different job characteristics, namely: job activity, location and provision of provident fund. S_j is the (real) salary offered for job j. One concern is that since cities have a higher cost of living than rural villages, positive compensating differentials for location might arise mechanically. To deal with this, we use the monthly Consumer Price Index (CPI)[12] to proxy for the cost of living and take the CPI value for the month in which the job ranking activity was implemented for the trainee. So, we

[12]Monthly CPI is obtained from the Ministry of Statistics and Program Implementation, Government of India for rural and urban areas at the state level and All-India level for our survey period. http://164.100.34.62:8080/cpiindex/Default1.aspx

deflate the salary for jobs in rural Uttar Pradesh by the monthly CPI of rural Uttar Pradesh; the salary for jobs in cities of Uttar Pradesh and Delhi by the monthly CPI for urban Uttar Pradesh and Delhi respectively and for jobs in the rest of the country, we deflate using the the All-India urban CPI for that month.

To calculate the compensating differentials, we then use the $\hat{\beta}$ and $\hat{\gamma}$ estimated in equation (1) above. Specifically, the ratio $-\hat{\beta}_k/\hat{\gamma}$ gives us the salary (in real terms) that would be needed to compensate a trainee to make her indifferent (i.e. have no change in the rank R_{ij}), if (all else equal) a job characteristic X^k was changed. Columns (2), (5), (8) of table 5 report the results for this ratio for the whole sample and then across males and females respectively. Lastly, to be able to interpret the magnitude of the compensating differential, we calculate it as a percentage of the salary (in real terms) in a baseline job i.e. a desk job, in the same district of the trainee's residence that offers no provident fund. Columns (3), (6) and (9) of table 5 report this percentage for the whole sample, males and females respectively.

As reported in column (1) of table 5, on average, trainees prefer in-state jobs and jobs with provident fund, and the latter is only statistically significant for men. However, it is worth noting that while men seem to be almost indifferent between desk jobs and active jobs (e.g. delivery boys) and between jobs in their local area of residence and jobs in bigger cities within the state, this is not true of women. The premium on desk jobs and local jobs is large (more than 15%) for women, though neither is statistically significant at conventional levels. Consistent with the stronger preference for staying local among women, the in-state premium is 54% for men and 136% for men relative to staying in their home district. This is what we would have expected given the social context of North India. Lastly, both men and women are willing to give up around 15% of their salary to be able to get a job with social security.

3.3 Are the two sets of preferences consistent?

In the above sections, we have described two methods (one based on a hypothetical exercise and the other based on choosing between real alternatives) that were used to elicit trainee preferences across different job characteristics. The question that we now turn to is whether these two sets of preferences are consistent. To do this, we take the list of 11 jobs that were ranked by the trainees in section 3.2. For each of these 11 jobs, we weight each characteristic of the job by the number of points that was allocated to that job characteristic by the trainee in the hypothetical exercise discussed in section 3.1. We can therefore produce a *hypothetical* ranking of the 11 jobs. We then compare how the *hypothetical* ranking for these 11 jobs compares with the *actual* ranking of those jobs by regressing the actual rank on the hypothetical rank with individual fixed effects. Table 6 reports the regression results. The hypothetical ranking exercise seems to be strongly predictive of the stated ranks, indicating that these two sets of preferences are consistent. This gives us some confidence that the preferences from the job rankings (which we will use in the rest of the paper) are reliable and for example, not the result of being confused about the preference elicitation exercise.[13]

4 Do managers know what they need to know?

As discussed in section 2.1, since a sizable amount of the financial compensation from the government is contingent on successful placement and retention of the job, placements are a priority for training institutes. Moreover, the manager of each training center is also the placement officer, responsible for matching trainees with firms for interviews and making sure that they get placed. In this section, we identify the particular matching friction that we emphasize in this paper: the fact that the placement managers do not necessarily know the preferences of the people that they are placing, and hence are likely to inefficiently match trainees to jobs.

To begin, we first examine if managers are aware of trainee preferences over jobs. To do this, we use the *same* list of 11 jobs that was provided to the trainees for ranking (in section 3.2) and for *each* trainee, ask managers to list (in order of preference) three jobs out of the 11 jobs that the trainee would like to work in. Using the manager and trainee preferences, we construct four measures of "how well" a manager knows her trainee's preferences.[14] As a benchmark, we can compare each of our measures (described below) to two hypothetical scenarios: one where the manager responds with a random list of jobs, and one where the manager has perfect knowledge of trainee preferences and responds based on that. The results for this activity are reported in figure 2 and table 7. We now discuss the four measures in detail below:

1. Measure #1: We consider a job that was picked by the manager as the best job for a trainee and report the rank provided by the trainee for that same job. If it were done randomly, the average rank should be close to 5.5 and if the manager knew the preferences of the trainee perfectly, this should be 11. In row (1) in table 7 we see that the average is 7.2 using actual trainee preferences. This does significantly better than a random process but significantly worse than the case where preferences were known perfectly.

2. Measure #2: We take all the three jobs chosen by the center manager and report the average rank given by the trainee for these jobs. This measure therefore gives us an idea of how good the manager is at knowing the preferences of the trainee on average. As reported in row (2) of table 7, random choice would have generated an average rank of approximately 6 while in the perfect information case it should be 10. The average observed in the data is 6.76, which again does better than a random allocation, but worse than perfect knowledge.

[13] It also suggests that hypothetical elicitation of job preferences can be indicative of the actual preferences, which is useful for future research.

[14] The results are very similar if we use the hypothetical preferences generated from a trainee's stated job priorities as described in section 3.1.

3. Measure #3: We take the highest rank assigned by the trainee to one of the three jobs picked for him by the center manager. Random choice would give us an average rank of 8.25 and if preferences were known perfectly by the manager, this should again be 11. But as reported in row (3), the average observed in the data is 9.38, which is again statistically better than a random process and worse than perfect information.

4. Measure #4: We consider the correlation between the rank orderings of the manager and the rank ordering of the trainee. With random choice, this correlation should be 0, while in the perfect information case, this correlation should be 1. With the job rankings, the average correlation is 0.1 in the data. This is better than the random process, but far worse than the perfect information case.

The above activity therefore identifies the friction that is at the heart of this paper: the placement manager, who is directly and completely responsible for the matching of job-seekers to jobs, does not seem to know the preferences of many of the job-seekers. Irrespective of which measure we use, she does do better than choosing jobs completely at random, but is nowhere near perfect information. Furthermore, as shown in figure 2, even across trainees, there seems to be a considerable amount of variation in the knowledge of manager. For example, in 20.5% of the cases, the manager is able to almost perfectly match the preferences of the trainee (correlation coefficient of 0.9 or more) while in 15.9% cases however, there is almost perfect negative correlation between the choices of the manager and those of the trainee (correlation coefficient of -0.9 or less).

5 The impact of informing managers

After eliciting preferences of trainees across jobs and establishing the manager's lack of knowledge of these preferences, we describe the randomized control trial associated with informing the centre managers about job preferences of the individuals who they are in charge of placing and the consequences it had.

5.1 Intervention details

In the placement week (which is the last week of the training program), the manager contacts various firms for job vacancies, and is therefore instrumental in matching trainees to these job interviews. The aim of our intervention was to reduce the asymmetry of information on trainee preferences over the set of firms as follows: trainees in each batch were randomized into two groups: for the first group, henceforth the Treatment group, we provided a description of the job characteristics for the top four jobs ranked by the trainee to the manager (see figure A3 for two examples). For the second group, henceforth the Control group, no trainee preferences were shared with the manager. As reported in panel A of table A4, the trainees in the two groups have similar observable characteristics. As reported in panel B, managers on average also have similar knowledge of preferences of trainees in the two groups, as captured by the four measures discussed in the previous section.

5.2 The impact on the number of interviews

We begin by examining whether the treatment had any effect on trainees getting more interviews or a different set of interviews. To examine this, we run the following specification with the results reported in columns (1)-(3) of table 8:

$$y_i = \alpha_b + \beta T_i + \gamma X_i + \varepsilon_i \tag{2}$$

where T_i is a dummy variable that takes the value 1 for if the trainee was in the treatment group and 0 for the control group. α_b is a batch/cohort fixed effect (since students were trained in batches of approximately 15 students each). X_i are a set of trainee characteristics like age, gender, education and dummy variables for if the trainee is currently a student and of lower caste.[15] y_i in column (1) of table 8 is the number of interviews received by trainee i, and in column (2), the number of interviews conditional on getting at least one. Column (3) and (4) are then dummy variables that equal 1 if the trainee gets an interview for (a) any job and (b) at least one job in her four most-preferred jobs respectively.[16] As reported in columns (1)-(3), there are no differential effects of the treatment on the number of interviews received by a trainee.

5.3 Quality of interviews: data challenges

Given that there is no effect on the number of interviews, it is somewhat easier to interpret the next set of results, which are about the quality of the match. We examine whether treated trainees were matched to interviews that they preferred more. There were two challenges that we encountered with the placement data: first, in the set of 11 jobs that were ranked by the trainees, we had varied the designation of the job (between active and desk jobs). However, most of the firms that candidates were actually matched to did not specify the type of job that they would place the trainee in, and so we could not match this dimension of preferences with the data. We therefore take the 11 jobs and average the rank over the designation dimension. This leaves us with 8 jobs for every trainee that now only vary in terms of salary, location and provident fund.

The bigger challenge was that if we took the complete set of combinations along the three dimensions (salary, location and provident fund) we would have 18 potential jobs. However,

[15] The results are robust to controlling for the trainee's average job rank for the three jobs chosen by the manager (measure #2 in section 4), to account for the fact that a manager may have more information on preferences for certain trainees.

[16] Details on interviews were collected in a follow up phone survey, where we were able to reach 293 out of the 338 trainees (a response rate of 87%).

as discussed earlier, to make the activity more realistic, we dropped some jobs based on the previous placement experience of Skills Academy. In the placement data however, we do encounter interviews where the set of job characteristics do not correspond to the jobs ranked by trainees. Out of a total of 217 interviews that we have in our data, we are able to perfectly match around two-thirds of the interviews (141 to be exact) with those in the job ranking list. For the remaining interviews, we do not have a match (and hence we do not know the preference of the trainee). Going forward, we only consider interviews where we know the trainee preferences.[17] The last row of table A4 shows that the number of interviews that we were able to match with preference rankings is not correlated by treatment assignment, as one would expect.

5.4 Impact on match quality

Since the intervention involved providing the manager with information on the four most preferred jobs of the trainee, we examine the impact of this intervention on two outcome variables: (i) a dummy variable on whether a trainee received at least one interview for her four most preferred jobs and (ii) the (normalized) average preference across all interviews received by the trainee.[18] We then re-estimate regression (2) and report the results in columns (4) and (5) table 8. We see that treatment trainees are 10.7 pp (48.4%) more likely to get an interview for one of their four most preferred jobs. More generally, the average placement quality, as measured by the average rank across interviews and normalized by the mean and standard deviation of the control group, is 0.26 standard deviations higher for the treatment trainees as compared to control ones.[19]

6 General equilibrium impacts

Given that in our experiment the treatment and control job-seekers were competing for the same pool of interviews, the experiment cannot directly tell us whether our intervention increased aggregate welfare. In particular, the intervention may have actually made things worse on average when one includes the control group. This is because we gave managers information on preferences of roughly half the trainees they had to assign to interviews, while saying nothing about the others. This can easily lead a manager to move to an allocation which is worse on average, and from one

[17] In an alternate exercise, we use machine learning to predict preferences for all interviews and redo our analysis using all interviews instead of just the ones where we have an exact match. Qualitatively, the results remain the same and are available upon request.

[18] We average the preference (rank reported by the trainee) across all interviews received by the trainee and normalize it by the mean and standard deviation of the control group.

[19] As discussed in section 4, since there is a lot of variation in the knowledge managers have about trainee preferences, we also examine the heterogeneity of the treatment along this dimension. We do not find any heterogeneous treatment effects.

that is in the core to one which is not.[20] To illustrate this with a simple example, consider three jobs: $1, 2, 3$ and three job-seekers: a, b, c. Let their preferences be: $\{(1P_a3P_a2), (1P_b2P_b3)(3P_c2P_c1)\}$. In the original allocation, the manager has some very noisy information about b's top preference and nothing else. Based on that, she chooses the allocation $\{a \to: 3; b \to 1; c \to 2\}$. b gets what the manager's best information says should be her top choice. Now suppose the manager is given very precise information about $a's$ preference and decides that she has no reason not to give a his top preference and then switches b to job 3, to generate the allocation $\{a \to 1; b \to 3; c \to 2\}$. This is not in the core (as c and b would like to swap). Moreover the number of job-seekers who have their second preference reduces by one, while the number of people with the top preference is still one.

An alternate experimental design, where we randomized information on trainee preferences at the batch-level instead of at the trainee level (as we do here), would not have been subject to this problem. The disadvantage of randomizing at the batch-level however, is that we would require more batches to get enough statistical power to detect the treatment effects. In appendix B, we show through simulations that detecting the treatment effect requires almost half the number of batches if we randomize across trainees than across batches. Our choice of the experimental design was therefore dictated by two factors: limited resource capacity (some of these areas are very rural and expensive to operate logistically and survey regularly), and the operational uncertainty in these areas, since batches were irregular (due to erratic and seasonal demand for training programs). This implied getting more batches in our sample and following up with those trainees was challenging.

To make progress however, we take a more theoretical approach. Our main challenge is that we need to predict the allocation of interviews between the treatment and control trainees in the absence of our intervention. For this, we need to model the manager's decision rule based on the observed allocation of interviews in treatment and control. This is what we discuss in section 6.1. Next, assuming that this rule is a reasonable approximation for how the manager actually decides, we can generate the counterfactual allocation for individuals in treatment and control in the absence of the intervention and hence examine the impact of our intervention after accounting for the general equilibrium effect by incorporating the control group as well. This is what we discuss in section 6.2.

6.1 The manager's interview allocation rule

There are three components to understanding how a manager would have allocated interviews in the absence of our intervention. First, we need to define the manager's information set i.e. her knowledge of trainee preferences both with and without the intervention. Second, we need to devise an algorithm to allocate the set of interviews across trainees (conditional on the manager's

[20] An allocation in the core is where two trainees cannot swap interviews with each other to make both of them better off.

information set) and lastly, we need to examine how the simulated allocation with the intervention compares to the actual allocation that we can empirically observe. We discuss each step in detail in this section.

Information sets of the manager

We begin by restricting the information set of the manager on trainee preferences. First, we consider a *complete information* case, where the manager knows trainee preferences as revealed in the job ranking exercise (from section 3.2). Second, we consider a *no information* case, and base the allocation of jobs on what the manager *thinks* are trainees preferences as reported to us by her (from section 4).[21] This is a reasonable benchmark for what a manager would do in the absence of our intervention or if she cannot process the information we gave her. Finally, we construct a *hybrid information* case, where the manager knows the revealed preferences from the job ranking exercise for the treatment group (since we gave her that information), but only has her guesses (that she reported to us) for the control group. This would be the right benchmark if the manager has fully processed all the information available to her after our treatment.

Allocation mechanism for interviews

To assign a decision-rule to the manager in allocating these interviews, we assume under each of the hypothesized information sets, the manager chooses allocations that are in the core (i.e. allocations where two trainees cannot swap interviews with each other to them better off). We can then compare the predicted allocations under each hypothesized information set, with the actual empirical allocation to choose an information set that is most likely used by the manager. However, before we proceed, there are several clarifications related to the allocation algorithm that are in order: first, we assume that the manager has no preferences over which trainee should get which interview.[22] Second, we implement the following algorithm to identify allocations in the core: trainees in a batch are arranged in a random order and then allowed to pick an interview from the set of available interviews. So for example, after the first trainee has picked an interview, then the next one gets to pick from the remaining ones, and so on. Third, for almost all batches, there are more trainees than interviews–so any allocation rule would have multiple allocations in the core. To take this into account, we run the algorithm 25,000 times, each time ordering trainees randomly within each batch to simulate the set of allocations. We can therefore calculate the probability that a trainee i is matched to an interview for job j (denoted by p_{ij}).[23] Fourth, we empirically observe a few individuals in our data (less than 10%) getting multiple interviews.

[21] We ignore any uncertainty that the manager may have around these preferences.

[22] The manager could for example act in the firm's interest and choose certain trainees because they fit the firm's needs better. That is ruled out by our assumption.

[23] Note that a 'job' in our setting is purely defined by the salary, location and availability of a provident fund. Variation in any other dimension (work timings for example) is not captured. As a result, we empirically observe some people getting multiple interviews for the "same" job. In such cases, we sum the probabilities across these jobs to calculate the probability that a trainee i is matched to any interview for job j.

As shown in figure A5, there 6 out of 21 batches, where more than 15% of trainees get multiple interviews. So unless we make further assumptions on how individuals can trade 'bundles' of interviews, we cannot perfectly compare the theoretical and empirical allocations (since in the simulated allocations, every individual gets only one interview). For our main results, we therefore drop these six batches, though we also show that our results are robust to either dropping all batches where at least one trainee gets multiple interviews, or on the other hand, including all batches.

Results

With the above caveats in mind, for *each* of the three information sets of the manager, we can generate a probability that an individual i is matched with an interview for job j, which we denote by p_{ij}. We then compare p_{ij} to the empirical allocation of interviews. To do this, we create a dummy variable (D_{ij}) that takes a value 1 if a trainee i gets an interview j and 0 otherwise. Pooling all the interviews and trainees, we calculate $E(D_{ij}|p_{ij})$, which is the expected probability of *empirically getting* an interview, conditional on the theoretical probability that a trainee *should* get one according to our allocation rule. Figure 3 plots this relationship.

If managers are allocating efficiently conditional on their information set, this should coincide with the 45 degree line. As can be seen in the first graph of figure 3, allocations under the hybrid information set do a better job at explaining the empirical allocations as compared to the other two cases.[24] Moreover, as shown in the second graph of figure 3, most trainees have relatively low values of p_{ij}, which is not surprising given the scarcity of jobs. However, it is precisely for those high p_{ij} jobs, where more information to the manager seems to be crucial in improving the allocation of interviews. This is intuitive, since it is for these jobs that being able to identify the small number of people who really want them creates a potential for a large welfare gain.

6.2 General equilibrium impact

Once we have a model of the manager's decision rule, we can use it to create a counterfactual allocation of interviews in the absence of the intervention. We can use this to then examine the impact of our intervention, *after* accounting for the reallocation of interviews between the control and treatment trainees. To begin, under the two "no information" and "hybrid information" sets, we simulate the allocation of interviews (using the algorithm and protocol as described above). We can therefore calculate the probability (under each scenario) that a treatment and control trainee receives a job of rank r (where a rank of 1 is least preferred and 8 is most preferred job). We report these distributions in figure 4. As can be seen from the graph, there is a clear increase in the probability that a treatment trainee gets interviews for a job that she prefers more. We then

[24] In figure A4, we show that this is robust to whether we drop all batches where at least one trainee receives multiple interviews, or on the other hand, include all batches.

formalize this graph by estimating the following regression specification:

$$p_{im} = \alpha_b + \beta T_i + \gamma \text{Hybrid}_m + \delta T_i \times \text{Hybrid}_m + \eta X_i + \varepsilon_{im} \tag{3}$$

Since our intervention provides the manager with information on the four most preferred jobs by a trainee, p_{im} is then the probability that a trainee i gets allocated an interview for at least one of her four most preferred job under an allocation rule $m \in \{\text{No info, Hybrid}\}$. T_i is an indicator variable for if the trainee is in the treatment or control group and X_i are the set of individual controls used in previous regressions. As reported in table 9, under the "no-information" set, there is no statistical difference in the probability that a control or a treatment trainee gets allocated an interview for their four most preferred jobs. On the other hand, under the hybrid information set, the treatment trainees are 9.5 pp (or 38%) more likely to be allocated an interview jobs that they prefer more (top four), while the control trainees are unaffected.

7 Impact on placements and employment

The above analysis is suggestive that the intervention did have an impact on improving the efficiency of the matching process. However, are trainees matched with more-preferred job interviews, also more likely to accept these jobs and retain them for longer? This is after all the outcome the government most cares about. We now examine the effect of our treatment on placement and employment outcomes in this section, which is possible since we observe: (a) the reported preference of a trainee for a job; (b) various placement and employment outcomes for every trainee-job pair.

7.1 Measuring placement and employment outcomes

For a trainee i and job j, we consider three outcomes related to interviews and offers– (i) at least one interview; (ii) at least one offer; (iii) whether an offer was accepted; and four outcomes related to job retention and employment– (i) whether the trainee was employed in the same job three and six months later and (ii) whether the trainee was employed in any job three and six months later.[25] Let us denote these outcomes by y_{ij}. We then use two ways to aggregate these numbers: the first is an *unweighted* index where for each individual, where we aggregate outcomes across all jobs to create an individual-specific placement and employment index denoted by y_i, which takes the value 1 if $\sum_j y_{ij} > 0$ and 0 otherwise.[26] Second we create a *preference-weighted* index, $y_i^{pw} = \sum_j r_{ij} y_{ij}$, where for each individual, we aggregate outcomes across all jobs *after weighting* them with the individual's ranking for that job (r_{ij}).

The preference weighted index therefore captures the idea that a trainee likes her placement

[25] We were able to survey 91% of our trainees after six months, resulting in a decrease in sample size.
[26] For the jobs that the trainee had ranked, but got no interview, we set all outcome variables to zero.

and employment outcomes better. This is potentially important, both from a welfare point of view, but also from the point of view of the efficiency of the labor market, since high turnover and low labor force attachment are both policy concerns in India. Liking the job you were placed in after training better may improve worker retention, both at the level of the employing firm as well as at the level of the labor market (i.e., if you enjoy the job you are placed in and hence perform well, it may be possible to move to another, perhaps even more desirable job).[27]

7.2 Impact on placement and employment outcomes

We now turn to discussing the effects of our intervention on employment and placement outcomes. We start with the *unweighted* placement and employment outcomes between the control and treatment trainees. In columns (2) and (3) of table 10, we report the mean of the outcome variable in the control and treatment group along with their difference. 50% of trainees in the control group received at least one interview, 36.4% received at least one offer and only 18% accepted an offer. The difference between the treatment and control groups, as reported in column (4), is both negligible in magnitude and statistically insignificant. In two rounds of follow up surveys, three and six months after initial placement, we find that 9 (12)% of control (treatment) trainees were employed in the same job three months later, while 25 (29)% of control (treatment) trainees were employed in any job. The difference between the two groups is large in proportional terms (33% and 16% respectively), but too imprecise to be statistically significant at conventional levels. Six months later, only 2 (1)% of control (treatment) trainees were employed in the same job that they were placed in. However, 19 (27)% of control (treatment) trainees were more likely to be employed in any job six months later. This translates into a 42% more likelihood of treatment trainees being employed six months later ($p = 0.14$).[28]

Given that there is no difference in the probability of getting an interview or taking a job, the preference weighted index mainly captures quality differences in the interviews and the jobs. As discussed before, for each individual i, we create a preference-weighted outcome y_i^{pw} that aggregates outcomes across jobs by weighting them by an individual's preference for it. To make comparisons between the control and treatment groups easier to interpret, we then normalize y_i^{pw} to have mean zero and standard deviation 1 for the control group trainees. Columns (5)-(7) of table 10 then report the mean of this normalized index for all the outcome variables. As can be seen, treatment trainees have a 0.27 standard deviation better quality of interviews, 0.19 standard deviation better quality offers and 0.24 standard deviation higher acceptance of these offers. All of these results are significant at conventional levels. These trainees are also 0.25 standard deviations more likely

[27] In our data, for trainees that got at least one interview, there is a positive (though statistically insignificant) correlation between the quality of initial placement and subsequent retention of that job for 3-6 months after placement.

[28] Similar to previous regression specifications, we estimate: $y_i = \alpha_b + \beta T_i + \gamma X_i + \varepsilon_i$, where y_i is the unweighted set of placement and employment outcomes. As reported panel A of table A5, the results do not change even after controlling for differences across trades, batches and individual characteristics.

to stay in this better quality job three months after placement ($p = 0.04$), though the effect does not persist six months after placement mainly because almost no one stays in the same job for six months.[29] In other words, better matching to interviews as a result of treatment does result in better employment outcomes, at least in the short-run.

8 Conclusion

This paper identifies an important potential source of mismatch in the Indian labor market – that intermediaries (placement managers in our context) who are responsible for matching job-seekers to jobs do not know the preferences of these job-seekers and therefore assign them to the "wrong" jobs. We provide evidence for this mismatch using the placement process for a large vocational training firm in India and examine the extent to which provision of information on preferences can lead to a better allocation of interviews, jobs and employee welfare. We see this paper as a part of a larger research agenda of understanding search costs and mismatch in the labor market and ways to reduce them. While others have emphasized externalities (Abel et al. (2016); Bassi et al. (2017); Pallais (2014)) and incentive problems (Krug and Stephan (2013); Behaghel et al. (2014); Laun and Thoursie (2014)), we show an example where the benefits are internal to the firm and the firm has strong incentives to get it right, but the outcome is nevertheless inefficient in the sense that some easily gathered information could lead to a much better allocation. In this sense, this is related to work by Bloom et al. (2013) in understanding the inefficient management practices in India. Understanding why managers do not use this information or at least trying to gather it would be the next logical step in this agenda.

Going beyond the specific issue of informational asymmetry, the question of how to get more of these trainees to stay in the labor market is clearly critical if a country like India is to be able to harvest its "demographic dividend". There is some hint that better matching can keep workers in the labor market (as reported in the previous section on job retention), but the effect while large is not statistically significant at conventional levels. Redoing our experiment or other interventions that improve matching with a bigger sample size is obviously one key step in either confirming this hypothesis or rejecting it. However, the broader result of the intervention improving job retention in the short term but not in the longer term resonates with conclusions drawn from other research studies across various countries in Africa (Abebe et al. (2017); Blattman et al. (2019)) and South-East Asia (Beam (2016)). This suggests the importance of understanding the drivers behind employment choices among the youth (such as aspirations, social norms, peer-effects etc.) better to be able to target policy interventions related to employment and job search more effectively.

Lastly, it may be important to start a culture of unpaid internships in firms for high school

[29] Similar to previous regression specifications, we estimate: $y_i^{pw} = \alpha_b + \beta T_i + \gamma X_i + \varepsilon_i$, where y_i^{pw} is the preference weighted set of placement and employment outcomes. As reported panel B of table A5, the results do not change even after controlling for differences across trades, batches and individual characteristics.

students so that they can learn what they like–the high quite rates that we see after placement, suggest that they often do not know what they are getting into. It is also important to try to persuade the youth to be more realistic about their employment options, possibly by engaging with social influencers and by highlighting the importance of getting started early.

References

Abebe, Girum, Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, Simon Quinn, and Forhad Shilpi, *Job fairs: Matching firms and workers in a field experiment in Ethiopia*, The World Bank, 2017.

Abel, Martin, Rulof Burger, and Patrizio Piraino, "The Value of Reference Letters-Experimental Evidence from South Africa," *Harvard University. Processed*, 2016.

Acemoglu, Daron, "A microfoundation for social increasing returns in human capital accumulation," *The Quarterly Journal of Economics*, 1996, *111* (3), 779–804.

_ , "Training and innovation in an imperfect labour market," *The Review of Economic Studies*, 1997, *64* (3), 445–464.

Banerjee, Abhijit V and Sandra Sequeira, "Spatial Mismatches and Imperfect Information in the Job Search," 2020.

Bassi, Vittorio, Aisha Nansamba, and BRAC Liberia, "Information frictions in the labor market: Evidence from a field experiment in uganda," *University College London. Processed*, 2017.

Beam, Emily A, "Do job fairs matter? experimental evidence on the impact of job-fair attendance," *Journal of Development Economics*, 2016, *120*, 32–40.

Behaghel, Luc, Bruno Crépon, and Marc Gurgand, "Private and public provision of counseling to job seekers: Evidence from a large controlled experiment," *American Economic Journal: Applied Economics*, 2014, *6* (4), 142–174.

Blattman, Christopher, Stefan Dercon, and Simon Franklin, "Impacts of industrial and entrepreneurial jobs on youth: 5-year experimental evidence on factory job offers and cash grants in Ethiopia," Technical Report, National Bureau of Economic Research 2019.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

Card, David, Jochen Kluve, and Andrea Weber, "Active labour market policy evaluations: A meta-analysis," *The economic journal*, 2010, *120* (548).

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora, "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment," *The quarterly journal of economics*, 2013, *128* (2), 531–580.

Dammert, Ana C, Jose Galdo, and Virgilio Galdo, "Integrating mobile phone technologies into labor-market intermediation: a multi-treatment experimental design," *IZA Journal of Labor & Development*, 2015, *4* (1), 11.

Diamond, Peter A, "Wage determination and efficiency in search equilibrium," *The Review of Economic Studies*, 1982, *49* (2), 217–227.

FICCI, "Reaping India's Promised demographic Dividend," 2013.

Franklin, Simon et al., "Location, search costs and youth unemployment: A randomized trial of transport subsidies in Ethiopia," *Centre for the Study of African Economies Working Paper WPS/2015-11*, 2015.

Groh, Matthew, David McKenzie, Nour Shammout, and Tara Vishwanath, "Testing the importance of search frictions and matching through a randomized experiment in Jordan," *IZA Journal of Labor Economics*, 2015, *4* (1), 7.

International Labour Organization, "Industrial Training Institutes in India; The Efficiency Study Report," 2003.

Krug, Gerhard and Gesine Stephan, "Is the contracting-out of intensive placement services more effective than provision by the PES? Evidence from a randomized field experiment," 2013.

Laun, Lisa and Peter Skogman Thoursie, "Does privatisation of vocational rehabilitation improve labour market opportunities? Evidence from a field experiment in Sweden," *Journal of health Economics*, 2014, *34*, 59–72.

Mortensen, Dale T and Christopher A Pissarides, "Job creation and job destruction in the theory of unemployment," *The review of economic studies*, 1994, *61* (3), 397–415.

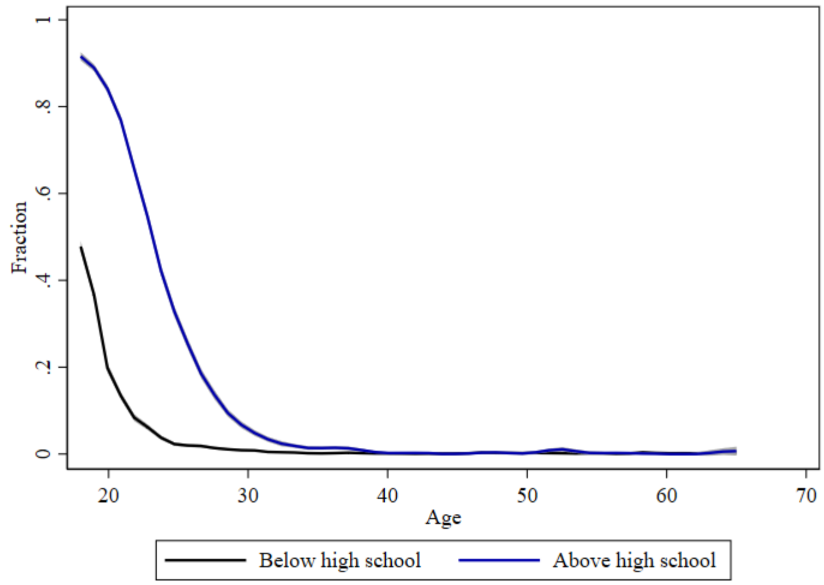
NSDC, "Overcoming India's Skills Challenge," 2013.

Pallais, Amanda, "Inefficient hiring in entry-level labor markets," *The American Economic Review*, 2014, *104* (11), 3565–3599.

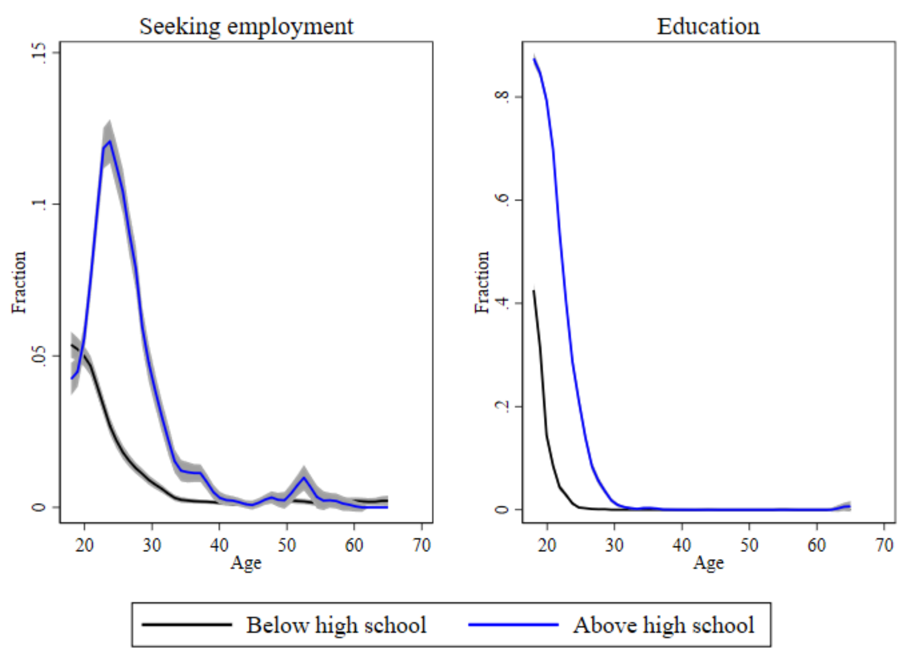
World Bank, "Skill Development in India: The Vocational Education and Training System.," *South Asia Human Development Sector Series*, 2008.

Figure 1: Non-employment rates by education status

(a) Whole sample

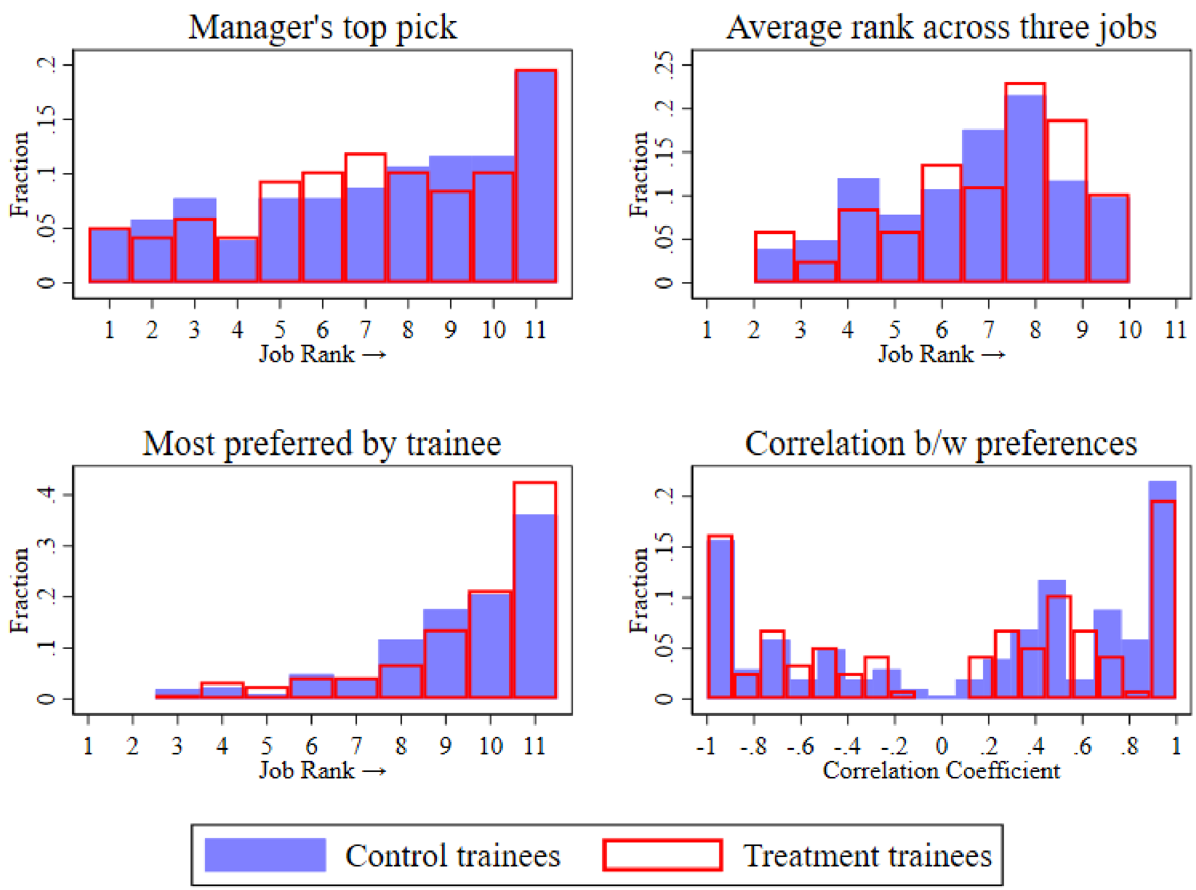


(b) Across seeking employment and education categories



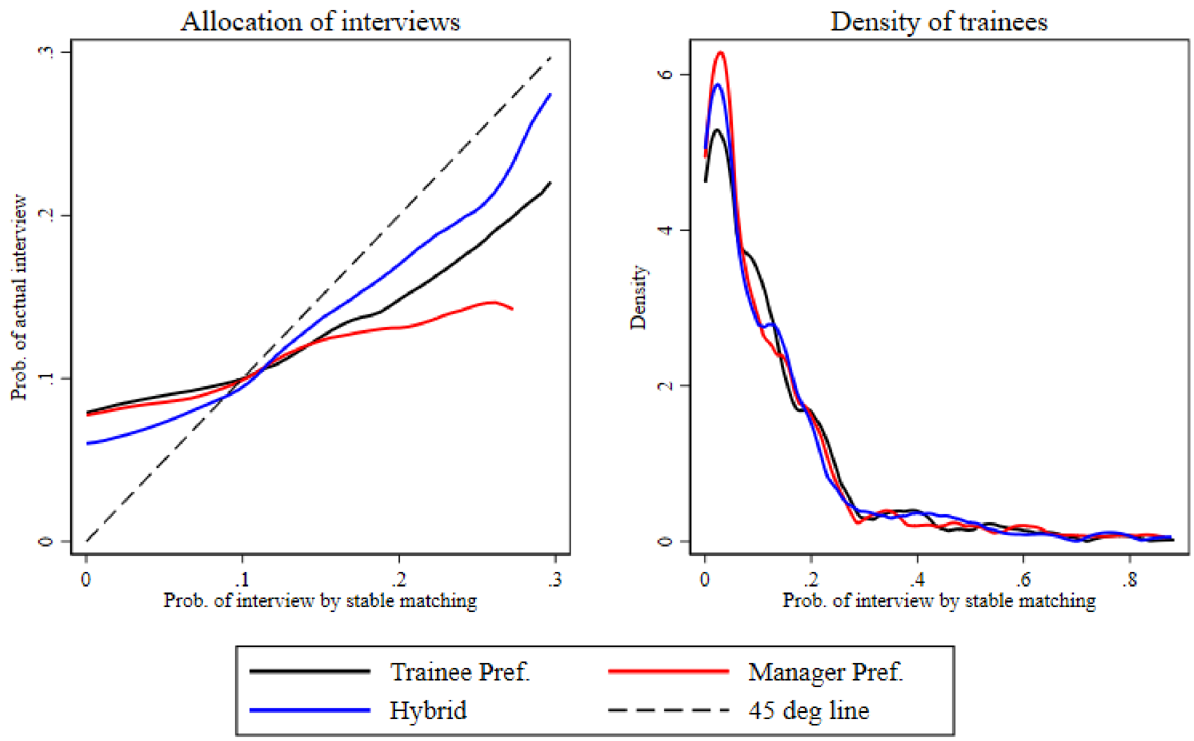
Notes: The above graph uses data from the 68th round of the National Sample Survey (NSS) for males between the age 18-65 years. The black line shows the non-employment rate for high-school dropouts, while the blue line reports it for high-school (and above) educated males. The first graph shows the non-employment rates for the whole sample, while the second graph breaks down the sample by those seeking employment (left) and those currently pursuing education (right).

Figure 2: Manager's knowledge of trainee job rankings



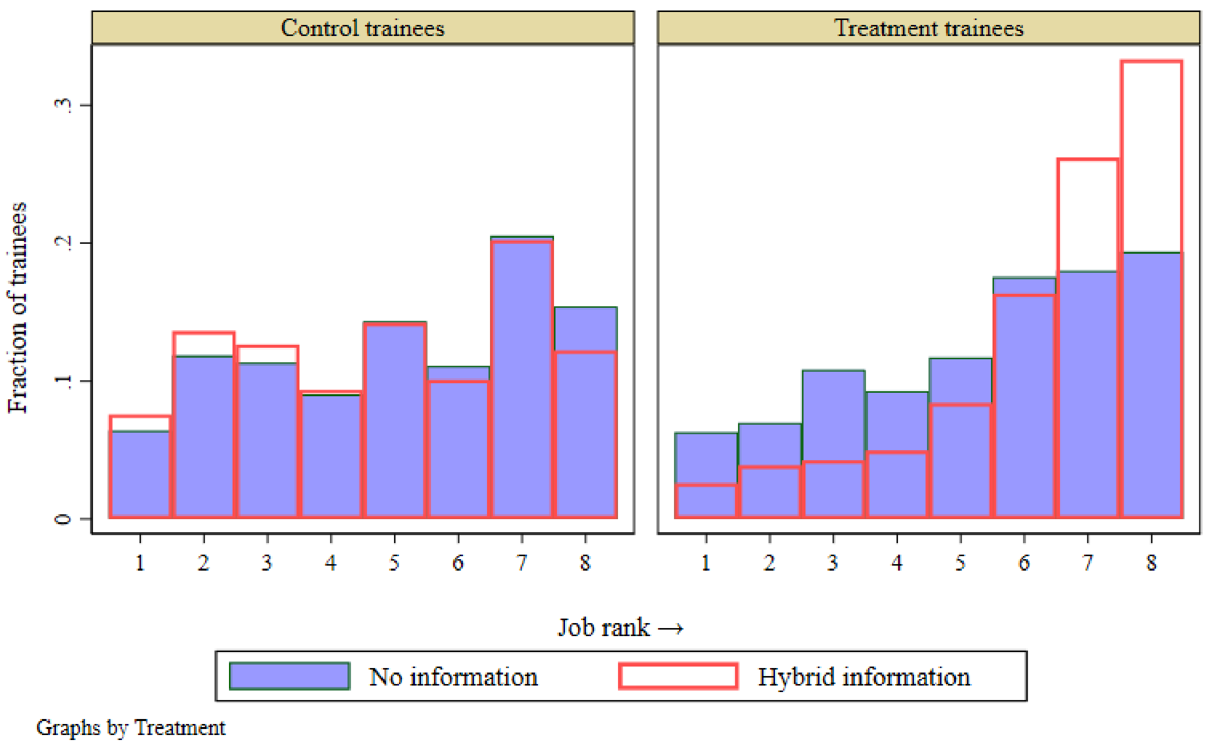
Notes: The above plots histograms for the four measures (discussed in the paper) of "how well" a manager knows the preferences of a trainee. Job rank goes from 1 to 11 where 11 is the most-preferred job by the trainee. We report the distribution separately for the control (blue) and treatment (red) trainees.

Figure 3: Simulated and empirical allocations



Notes: The horizontal axis in both graphs show the probability that a trainee i gets an interview for job j as allocated by the algorithm described in the paper. The first graph on the left compares the simulated allocation to the empirical allocation under the three information sets of the manager, depicted by the black, red and blue lines for the full information, no information and hybrid case. The dotted line is the 45 degree line. The second graph on the right shows the density of trainees across the simulated probability distribution.

Figure 4: Probability of getting a job with rank r



Notes: The graph shows the probability that a trainee gets an interview for a job rank with r, where 1 is the least preferred job and 8 is the most preferred job. The histogram then shows the fraction of trainees (control group on the first graph on the left, treatment group on the second graph on the right) who get a job of rank r. The blue bars show the distribution under the no information case, while the red bars show the distribution under the hybrid information case.

Table 1: Description of the sample of trainees

	Study Sample	NSS Sample: All India	NSS Sample: Rural U.P. and Delhi
	(1)	(2)	(3)
Female	0.48	0.44*	0.43**
Age	20.92	25.37***	24.59***
Married	0.11	0.46***	0.52***
Education (years)	13.78	13.49***	13.54***
HH Size	5.22	5.39	7.11***
Hindu	0.93	0.76***	0.92
Caste (General)	0.26	0.42***	0.42***
Caste (OBC)	0.37	0.37	0.41*
Caste (SC)	0.37	0.11***	0.15***

Notes: Column (1) reports the mean for the study sample. This is compared to the 68th round of the National Sample Survey in columns (2) and (3). The NSS sample is constrained to individuals with at least high school level of education and between the age groups of 18-35 years of age to match the eligibility of the study sample. Column (2) reports the mean in the NSS sample for the whole of India, while column (3) reports the mean in the NSS sample for rural Uttar Pradesh and Delhi only. Asteriks report the results from a t-test that compare the means in columns (2) and (3) to the mean in column (1). Female takes the value 1 if the individual is female and 0 otherwise. Married is a dummy that takes the value 1 if married and 0 otherwise. Education and age are reported in years. Hindu is a dummy that takes the value 1 if the individual is a Hindu and 0 otherwise. Caste variables are also dummies that take the value 1 if the individual belongs to that caste and 0 otherwise. * $p< 0.1$, ** $p<0.05$ and *** $p< 0.01$ level of significance.

Table 2: Labor market aspirations

	N	Mean	S.D.
	(1)	(2)	(3)
Panel A: Sectors for employment			
Healthcare	538	0.32	0.47
Banking	538	0.25	0.43
Retail	538	0.15	0.36
Hospitality	538	0.09	0.29
IT	538	0.08	0.27
BPO	538	0.04	0.21
Panel B: Salary and social security			
Salary (in rupees)	370	15036.49	9550.43
Provident Fund	370	0.98	0.13
Prefer public sector job?	370	0.96	0.18
Panel C: Location preferences			
	Location of job		
Respondent Residence	Residence area	City in Uttar Pradesh	Rest of India
Rural UP (N = 297)	0.18	0.74	0.08
Delhi (N = 67)	0.97	-	0.03

Notes: Panel A reports the means from a dummy variable that takes a value 1 if the individual ranks that sector as his/her most preferred sector of employment and 0 otherwise. Salary is the monthly salary reported in Indian rupees. Provident Fund and Prefer public sector job are dummy variables that take the value 0 if no and 1 if yes. Panel C reports job location preferences conditional on the residence of the trainee.

Table 3: Distribution of 100 points

Job characteristic	N	Whole sample Mean	S.D.	Male	Female	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Salary	538	26.11	18.20	26.63	25.55	0.49
Location	538	18.67	15.70	16.59	20.91	0.00
Designation	538	19.02	16.17	20.05	17.9	0.12
Nature of work	538	10.16	11.67	10.25	10.06	0.85
Job security	538	13.35	15.96	13.33	13.37	0.98
Social status	538	12.70	15.34	13.15	12.21	0.48

Notes: Columns (2) and (3) report the mean and standard deviation of the average points given to the job characteristic. Columns (4) and (5) report the average points given to the job characteristic by males and females respectively. Lastly, column (6) reports the p-value of a t-test that tests the statistical difference between columns (4) and (5).

Table 4: Job ranking and strategic reporting

		Percent trainees who ranked job in			Salience of job ranking		
	N	Bottom three jobs	Rank 4-8 jobs	Top three jobs	Low salience	High salience	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Job 1	338	0.46	0.4	0.14	4.73	4.38	0.3
Job 2	338	0.38	0.42	0.2	5.45	5.08	0.31
Job 3	338	0.33	0.44	0.23	5.38	5.55	0.62
Job 4	338	0.31	0.49	0.2	5.43	5.61	0.59
Job 5	338	0.12	0.54	0.33	7.05	6.52	0.08
Job 6	338	0.18	0.5	0.31	6.54	6.66	0.72
Job 7	338	0.13	0.38	0.49	7.75	7.71	0.9
Job 8	338	0.32	0.47	0.21	5.31	5.6	0.38
Job 9	338	0.39	0.42	0.19	4.84	5.15	0.35
Job 10	338	0.19	0.49	0.32	6.39	6.53	0.69
Job 11	289	0.19	0.39	0.42	6.7	7.32	0.11

Notes: Columns (2)-(4) report the fraction of trainees who ranked a job amongst the bottom three, rank 6-8 and top 3 jobs. Columns (5) and (6) report the average rank that is given to a job by the trainee in the low and high salience groups. A higher rank indicates more preference. Column (7) reports the p-value of a t-test that tests the statistical difference between columns (5) and (6).

Table 5: Preferences for job characteristics

	Whole sample			Male			Female		
	$\hat{\beta}_k$	$\frac{-\hat{\beta}_k}{\hat{\gamma}}$	Percent of salary	$\hat{\beta}_k$	$\frac{-\hat{\beta}_k}{\hat{\gamma}}$	Percent of salary	$\hat{\beta}_k$	$\frac{-\hat{\beta}_k}{\hat{\gamma}}$	Percent of salary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Active	-0.132	1.89	5.46	-0.0505	0.50	1.45	-0.233	5.18	15.05
	(0.149)			(0.202)			(0.216)		
Same state	-0.0621	0.89	2.57	-0.0254	0.25	0.73	-0.289	6.42	18.66
	(0.198)			(0.279)			(0.278)		
Out of state	-1.855***	26.50	76.77	-1.904***	18.85	54.48	-2.103***	46.73	135.81
	(0.304)			(0.416)			(0.442)		
PF	0.368**	-5.26	-15.23	0.549***	-5.44	-15.71	0.220	-4.89	-14.21
	(0.114)			(0.144)			(0.177)		
Salary (Real)	0.0700***	-1	-	0.101***	-1	-	0.0450***	-1	-
	(0.00701)			(0.00962)			(0.00960)		
Real salary for desk job, same dist., no PF		34.52			34.6			34.41	
N		3669			1919			1750	
R^2		0.112			0.182			0.094	
Trainee FE		Yes			Yes			Yes	

Notes: Salary is reported in real terms. Columns (2), (5), (8) report the compensating differential in real rupees and in columns (3), (6), (9) as a fraction of the real salary for a desk job in the same district without PF. Standard errors are clustered at the trainee level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 6: Hypothetical and actual preferences

	Reported Rank	
	(1)	(2)
Hypothetical Rank	0.145*	0.145***
	(0.0261)	(0.0259)
N	3647	3658
R^2	0.032	0.057
Individual Controls	Yes	No
Centre FE	Yes	No
Trade FE	Yes	No
Individual FE	No	Yes

Notes: Reported rank is the rank given by a trainee in the job ranking exercise. Column (1) includes individual controls of age, gender, years of education, religion, caste and whether the trainee has any work experience or not along with center and trade fixed effects. Column (2) reports results using individual fixed effects instead. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 7: Manager knowledge of trainee preferences

	Measure of knowledge	Reported rank	Random Process	Perfect Knowledge
	(1)	(2)	(3)	(4)
1.	Rank of manager's top choice	7.2^{***}_{***}	5.5	11
2.	Average Rank by trainee	6.76^{***}_{***}	6	10
3.	Most preferred by trainee	9.38^{***}_{***}	8.25	11
4.	Correlation b/w preferences	0.1^{**}_{***}	0	1

Notes: Each row in column (1) is a different measure of the manager's knowledge of trainee preferences with the measure explained in the heading. Column (2) reports the average job rank as reported in the job choice exercise. Column (3) calculates the rank as if this process was done randomly. Column (4) calculates the rank as if the managers had perfect knowledge of trainee preferences. The asteriks in the top and bottom row are the results from a t-test that compares the value to column (3) and (4) respectively. * $p<0.1$, ** $p<0.05$ and *** $p<0.01$ respectively.

Table 8: Impact on interviews and job characteristics

	No. of interviews		At least one interview		Normalized job preference
	Unconditional	Conditional	Any job	Top-four job	
	(1)	(2)	(3)	(4)	(5)
Treatment	0.0900	0.135	0.0188	0.107**	0.262**
	(0.0906)	(0.109)	(0.0529)	(0.0526)	(0.111)
N	293	149	293	293	293
R^2	0.330	0.388	0.253	0.256	0.184
Control mean	0.693	1.386	0.500	0.221	0.00
Ind. Controls	Yes	Yes	Yes	Yes	Yes
Batch FE	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1) and (2) report the number of interviews received by a trainee and the number of interviews conditional on receiving at least one respectively. Columns (3) and (4) create a dummy variable that takes a value 1 if a trainee receives at least one interview or an interview for a top-four preferred job respectively, and 0 otherwise. Job preferences in column (5) have been averaged across all interviews received by a trainee and then normalized by the mean and standard deviation in the control group. Individual controls used are the number of interviews, age, gender, years of education, indicator variables for whether the trainee is a student or not and whether from a SC/ST/OBC caste category. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 9: Probability of getting a top-four job

Depvar:	Prob. getting top-four preferred job	
	(1)	(2)
No info x Treat	0.0361	0.0224
	(0.0317)	(0.0191)
Hybrid x Control	-0.0191	-0.0191
	(0.0313)	(0.0220)
Hybrid x Treat	0.0952**	0.0952***
	(0.0449)	(0.0259)
Control, No Info. Mean	0.25	0.25
R^2	0.033	0.707
N	586	586
Ind. controls	No	Yes
Batch FE	No	Yes

Notes: The dependent variable is the probability that a trainee i gets an interview for at least one top-four job. No-info is a dummy variable that takes the value 1 if the allocation was simulated under the no-information case, while Hybrid is a dummy variable that takes the value 1 if the allocation was simulated under hybrid information case. Column (1) does not have any individual controls or batch FE while column (2) adds them. Individual controls used are the number of interviews, age, gender, years of education and manager knowledge of trainee preferences (measure #2), indicator variables for whether the trainee is a student or not and whether from a SC/ST/OBC caste category. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 10: Impact on job choice and employment outcomes

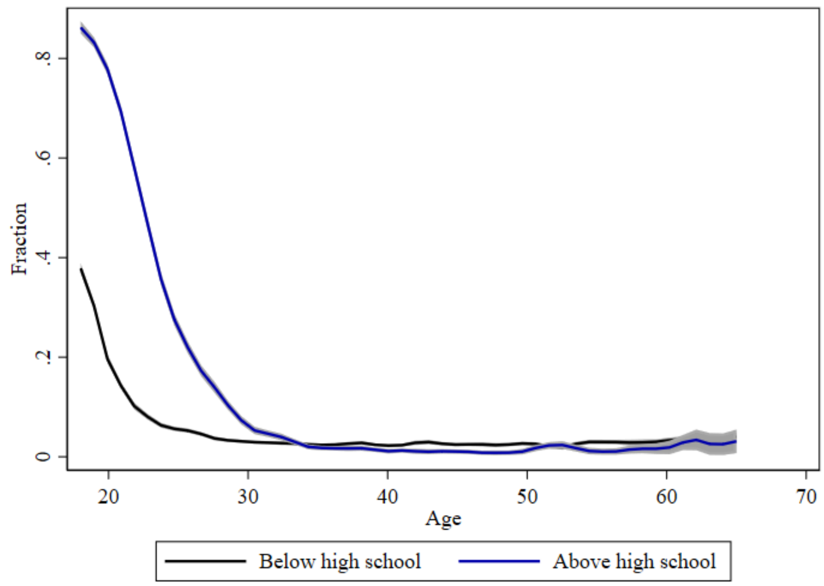
		Unweighted				Quality weighted	
		N	Control	Treatment	Diff.	Control	Treatment
		(1)	(2)	(3)	(4)	(5)	(6)
Placement:	Interview	293	0.5	0.52	0.02	0	0.27**
(Atleast one)			(0.5)	(0.5)		(1)	(0.95)
	Offer	293	0.36	0.36	0.00	0	0.19*
			(0.48)	(0.48)		(1)	(0.93)
	Accepted	293	0.18	0.18	0.00	0	0.24**
			(0.38)	(0.38)		(1)	(1.06)
Same jobs:	3 months	293	0.09	0.12	0.03	0.00	0.25**
			(0.28)	(0.32)		(1)	(1.09)
	6 months	266	0.02	0.01	-0.01	0	0.08
			(0.12)	(0.09)		(1)	(0.00)
Employed:	3 months	293	0.25	0.29	0.04		
			(0.43)	(0.45)			
	6 months	266	0.19	0.27	0.08		
			(0.4)	(0.45)			

Notes: Columns (2)-(4) report the fraction of various placement and job outcomes between control and treatment groups. Columns (5)-(6) report the same outcomes but weighted by trainee preferences and normalized to have mean 0 and standard deviation 1 for the control group. Difference in columns (4) is the difference between the treatment and control means. Standard deviations are reported in parentheses. * $p< 0.1$, ** $p<0.05$ and *** $p< 0.01$ level of significance.

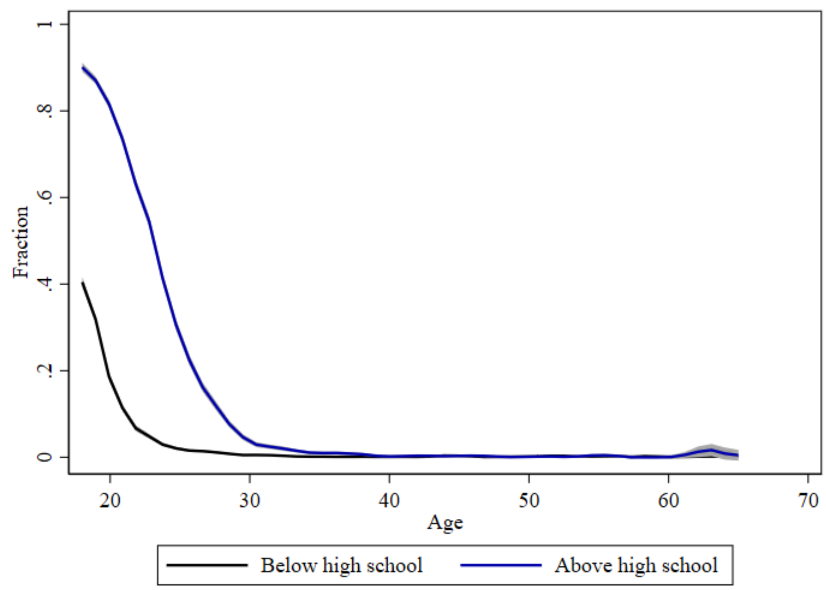
A Appendix figures and tables

Figure A1: Non-employment rates by education levels for NSS Rounds 64 and 66

(a) NSS Round 64 (2007)

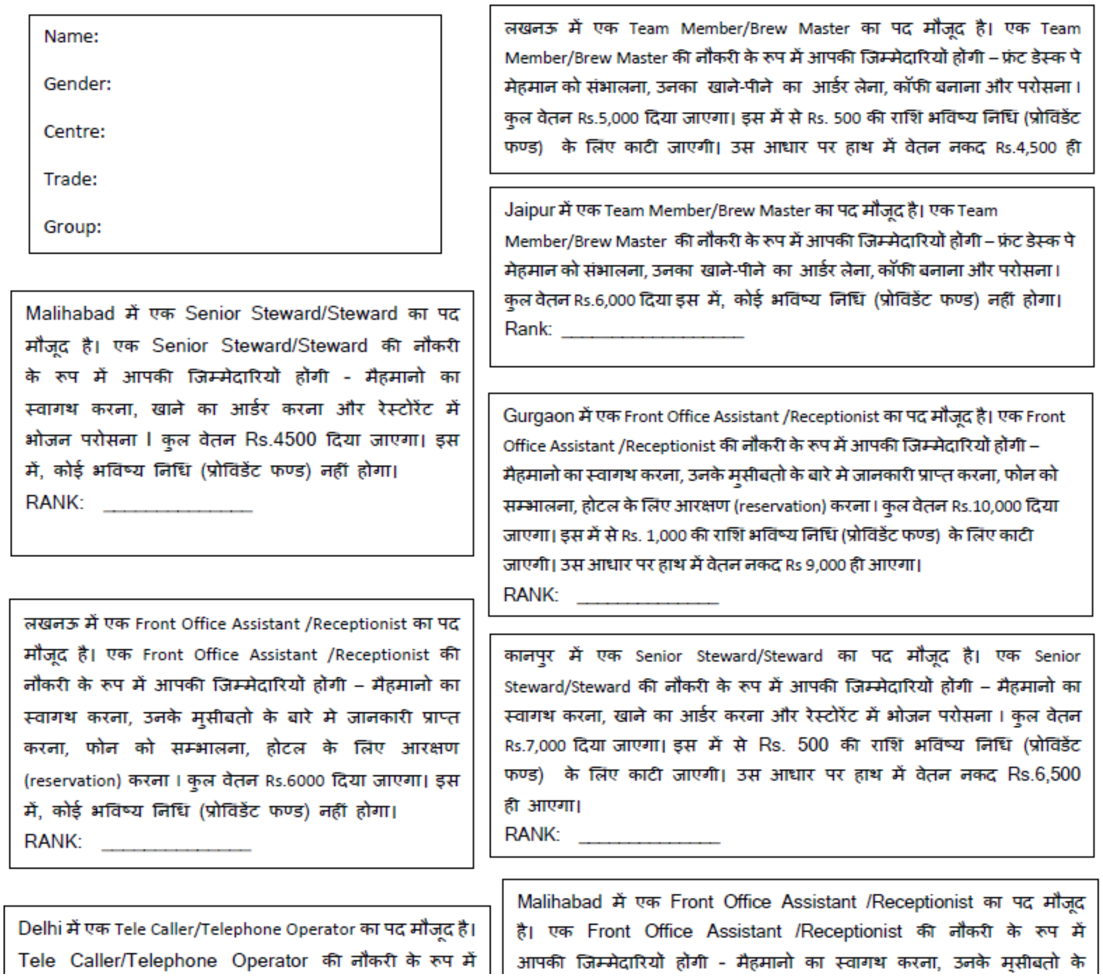


(b) NSS Round 66 (2009)



Notes: The above graphs use data from the 64th and 66th rounds of the National Sample Survey (NSS) for males between the age 18-65 years. The black line shows the non-employment rate for high-school dropouts, while the blue line reports it for high-school (and above) educated males.

Figure A2: Example of a job list


Name:

Gender:

Centre:

Trade:

Group:

Malihabad में एक Senior Steward/Steward का पद मौजूद है। एक Senior Steward/Steward की नौकरी के रूप में आपकी जिम्मेदारियों होगी - मेहमानो का स्वागथ करना, खाने का आर्डर करना और रेस्टोरेंट में भोजन परोसना । कुल वेतन Rs.4500 दिया जाएगा। इस में, कोई भविष्य निधि (प्रोविडेंट फण्ड) नहीं होगा।
RANK: ______________

लखनऊ में एक Front Office Assistant /Receptionist का पद मौजूद है। एक Front Office Assistant /Receptionist की नौकरी के रूप में आपकी जिम्मेदारियों होगी – मेहमानो का स्वागथ करना, उनके मुसीबतो के बारे मे जानकारी प्राप्त करना, फोन को सम्भालना, होटल के लिए आरक्षण (reservation) करना । कुल वेतन Rs.6000 दिया जाएगा। इस में, कोई भविष्य निधि (प्रोविडेंट फण्ड) नहीं होगा।
RANK: ______________

Delhi में एक Tele Caller/Telephone Operator का पद मौजूद है। Tele Caller/Telephone Operator की नौकरी के रूप में

लखनऊ में एक Team Member/Brew Master का पद मौजूद है। एक Team Member/Brew Master की नौकरी के रूप में आपकी जिम्मेदारियों होगी – फ्रंट डेस्क पे मेहमान को संभालना, उनका खाने-पीने का आर्डर लेना, कॉफी बनाना और परोसना। कुल वेतन Rs.5,000 दिया जाएगा। इस में से Rs. 500 की राशि भविष्य निधि (प्रोविडेंट फण्ड) के लिए काटी जाएगी। उस आधार पर हाथ में वेतन नकद Rs.4,500 ही

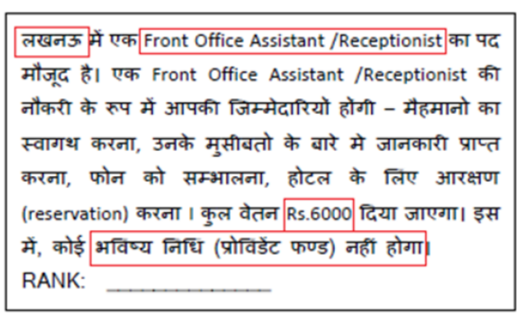
Jaipur में एक Team Member/Brew Master का पद मौजूद है। एक Team Member/Brew Master की नौकरी के रूप में आपकी जिम्मेदारियों होगी – फ्रंट डेस्क पे मेहमान को संभालना, उनका खाने-पीने का आर्डर लेना, कॉफी बनाना और परोसना। कुल वेतन Rs.6,000 दिया इस में, कोई भविष्य निधि (प्रोविडेंट फण्ड) नहीं होगा।
Rank: ________________

Gurgaon में एक Front Office Assistant /Receptionist का पद मौजूद है। एक Front Office Assistant /Receptionist की नौकरी के रूप में आपकी जिम्मेदारियों होगी – मेहमानो का स्वागथ करना, उनके मुसीबतो के बारे मे जानकारी प्राप्त करना, फोन को सम्भालना, होटल के लिए आरक्षण (reservation) करना । कुल वेतन Rs.10,000 दिया जाएगा। इस में से Rs. 1,000 की राशि भविष्य निधि (प्रोविडेंट फण्ड) के लिए काटी जाएगी। उस आधार पर हाथ में वेतन नकद Rs 9,000 ही आएगा।
RANK: ______________

कानपुर में एक Senior Steward/Steward का पद मौजूद है। एक Senior Steward/Steward की नौकरी के रूप में आपकी जिम्मेदारियों होगी – मेहमानो का स्वागथ करना, खाने का आर्डर करना और रेस्टोरेंट में भोजन परोसना । कुल वेतन Rs.7,000 दिया जाएगा। इस में से Rs. 500 की राशि भविष्य निधि (प्रोविडेंट फण्ड) के लिए काटी जाएगी। उस आधार पर हाथ में वेतन नकद Rs.6,500 ही आएगा।
RANK: ______________

Malihabad में एक Front Office Assistant /Receptionist का पद मौजूद है। एक Front Office Assistant /Receptionist की नौकरी के रूप में आपकी जिम्मेदारियों होगी - मेहमानो का स्वागथ करना, उनके मुसीबतो के

(a) Job list for ranking


लखनऊ में एक Front Office Assistant /Receptionist का पद मौजूद है। एक Front Office Assistant /Receptionist की नौकरी के रूप में आपकी जिम्मेदारियों होगी – मेहमानो का स्वागथ करना, उनके मुसीबतो के बारे मे जानकारी प्राप्त करना, फोन को सम्भालना, होटल के लिए आरक्षण (reservation) करना । कुल वेतन Rs.6000 दिया जाएगा। इस में, कोई भविष्य निधि (प्रोविडेंट फण्ड) नहीं होगा।
RANK: ______________

Location:	Lucknow
Designation:	Office Assistant/Receptionist
Salary:	Rs. 6000
Provident Fund:	Not provided

(b) Example of a job

Notes: Figure (a) shows an example of a job list that was provided to trainees while figure (b) provides example of one such job.

Figure A3: Example of two job lists given to the manager

Student Name : [redacted]

Job #1: Location = Gurgaon; Role = Delivery Boy; Salary = Rs. 9,000; Provident Fund = Yes (Rs. 500)

Job #2: Location = Gurgaon; Role = Receptionist; Salary = Rs. 8,500; Provident Fund = No

Job #3: Location = Jaipur; Role = Customer Sales Associate; Salary = Rs. 8,000; Provident Fund = Yes (Rs. 1,000)

Job #4: Location = Lucknow; Role = Customer Sales Associate; Salary = Rs. 8,000; Provident Fund = Yes (Rs. 1,000)

Student Name:[redacted]

Job #1: Location = Lucknow; Role = Customer Sales Associate; Salary = Rs. 8,000; Provident Fund = Yes (Rs. 1,000)

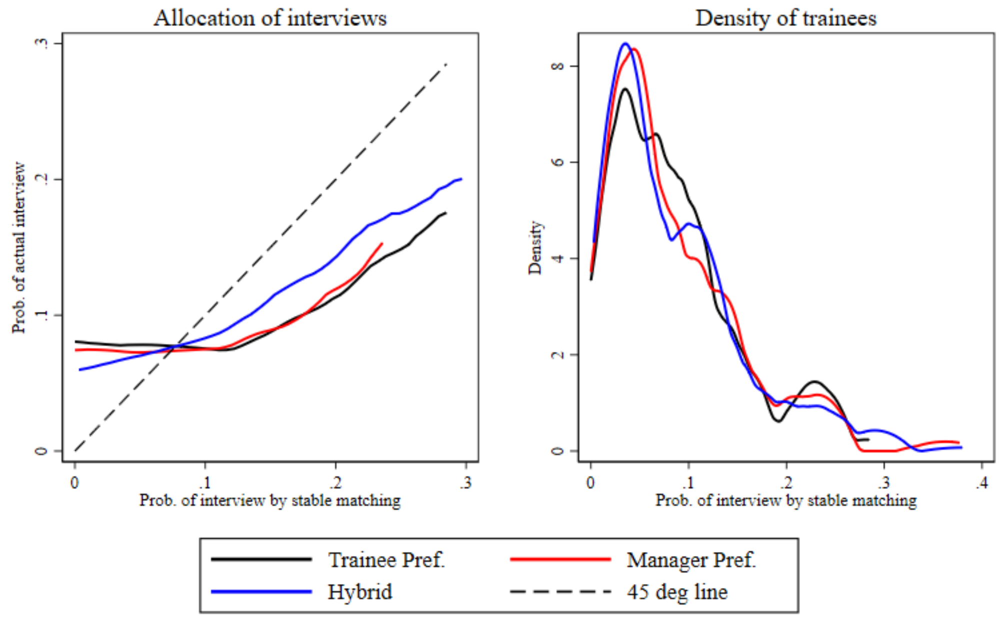
Job #2: Location = Lucknow; Role = Receptionist; Salary = Rs. 4,500; Provident Fund = No

Job #3: Location = Lambhua; Role = Customer Care Executive; Salary = Rs. 4,000; Provident Fund = No

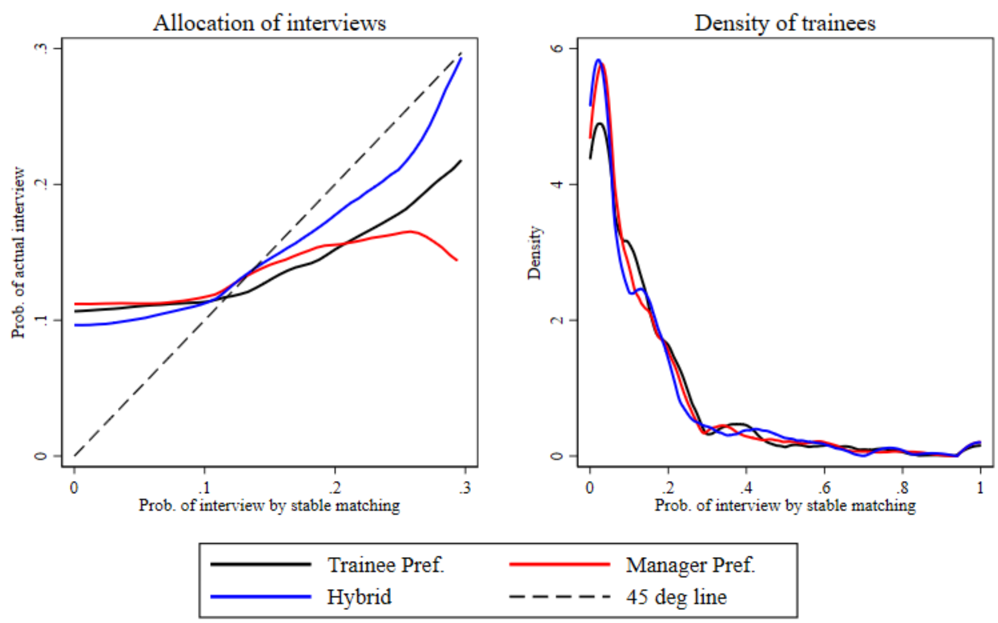
Job #4: Location = Lambhua; Role = Customer Sales Associate; Salary = Rs. 4,500; Provident Fund = Yes (Rs. 500)

Notes: The above figure provides an example of two job preference rankings for two trainees that were provided to the placement manager.

Figure A4: Simulated and empirical allocations



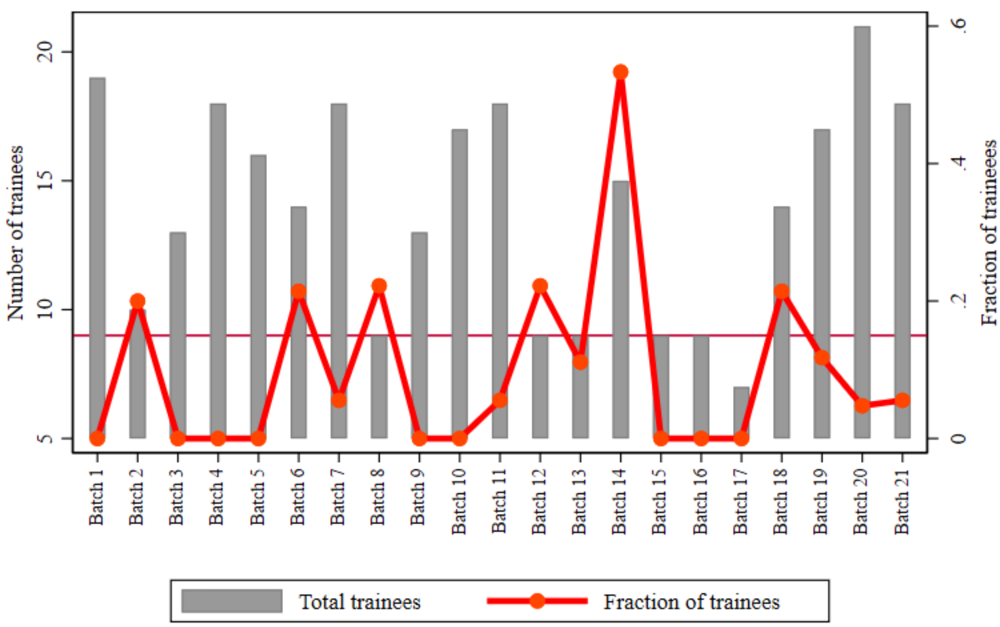
(a) Batches with no multiple interviews



(b) All batches

Notes: The horizontal axis in both graphs show the probability that a trainee i gets an interview for job j as allocated by the algorithm described in the paper. The graph on the left compares the simulated allocation to the empirical allocation under the three information sets of the manager, depicted by the black, red and blue lines for the full information, no information and hybrid case. The dotted line is the 45 degree line. The second graph on the right shows the density of trainees across the simulated probability distribution. Figure (a) shows the allocation of interviews with the sample restricted to only those batches where no trainee got multiple interviews. Figure (b) shows the allocation of interviews with all batches.

Figure A5: Distribution of trainees and interviews



Notes: The above figure shows the number of trainees (grey bars) in each of the 21 batches in our sample. The red line shows the fraction of trainees in each batch that got multiple interviews.

Table A1: Variation in job characteristics

Sr. No.	Job characteristic	Variation
1.	Salary	Low, medium or high
2.	Location	Local area of residence Within the state Outside the state in the rest of India
3.	Social security	No or Yes
4.	Designation	Desk/phone or activity intensive job

Table A2: Selection into job ranking activity

	N	Absent	Present	p-value
	(1)	(2)	(3)	(4)
Female	538	0.49	0.48	0.76
Age	538	21.11	20.80	0.23
Hindu	538	0.96	0.91	0.02
Caste (General)	538	0.26	0.26	0.83
Caste (OBC)	538	0.4	0.35	0.27
Caste (SC)	538	0.34	0.38	0.43
Education (years)	538	13.87	13.72	0.3
Work experience (years)	537	0.17	0.19	0.73
Father's age	447	50.3	49.41	0.28
Mother's age	486	45.1	44.85	0.72
Father education	442	8	7.98	0.97
Mother education	485	3.68	3.51	0.7

Notes: Columns (2) and (3) report the average values for a characteristic for trainees who were absent and present for the job ranking activity respectively. Column (4) reports the p-value of a t-test that tests the statistical difference between columns (2) and (3).

Table A3: Balance check for job ranking activity

	N	Low likelihood	High likelihood	p-value
	(1)	(2)	(3)	(4)
Female	338	0.49	0.46	0.52
Age	338	21.08	20.53	0.08
Hindu	338	0.9	0.93	0.32
Caste (General)	338	0.28	0.25	0.5
Caste (OBC)	338	0.35	0.35	0.97
Caste (SC)	338	0.37	0.39	0.72
Education (years)	338	13.67	13.78	0.49
Work experience (years)	337	0.18	0.19	0.91
Father's age	285	49.76	49.08	0.48
Mother's age	309	45.41	44.26	0.18
Father education	284	8.01	7.94	0.91
Mother education	309	3.53	3.49	0.94

Notes: Columns (2) and (3) report the average values for a characteristic for trainees who were assigned to the low and high likelihood groups for the job ranking activity respectively. Column (4) reports the p-value of a t-test that tests the statistical difference between columns (2) and (3).

Table A4: Balance check for the intervention

	N	Control	Treatment	p-value
	(1)	(2)	(3)	(4)
Panel A: Trainee characteristics				
Female	310	0.43	0.46	0.56
Age	310	20.88	20.70	0.59
Hindu	310	0.92	0.90	0.54
Caste (General)	310	0.24	0.25	0.79
Caste (OBC)	310	0.37	0.33	0.47
Caste (SC)	310	0.38	0.41	0.63
Education (years)	310	13.83	13.69	0.42
Work experience (years)	309	0.22	0.17	0.31
Father's age	266	50.41	48.80	0.11
Mother's age	287	45.66	44.17	0.09
Father education	263	8.52	7.53	0.11
Mother education	285	3.27	3.51	0.67
Panel B: Measures of manager knowledge of trainee preferences				
Measure #1	219	6.3	6.92	0.12
Measure #2	219	9.33	9.46	0.63
Measure #3	219	6.64	6.89	0.35
Measure #4	219	0.15	0.07	0.39
Exact job matches	217	0.61	0.68	0.25

Notes: Columns (2) and (3) report the average values for a characteristic for trainees who were assigned to the control and treatment groups where treatment group preferences on jobs ranked by the trainee were provided to the manager. Column (4) reports the p-value of a t-test that tests the statistical difference between columns (2) and (3). Panel A reports the observable characteristics of the trainees and panel B reports the four measures of manager knowledge of trainee preferences.

Table A5: Impact on job choice and employment outcomes

	Atleast one interview	Offer received	Offer accepted	Same job (3 months)	Employed (3 months)	Employed (6 months)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Unweighted placement outcomes						
Treatment	0.0146	-0.00354	-0.00201	0.0279	0.0335	0.0744
	(0.0528)	(0.0522)	(0.0437)	(0.0346)	(0.0521)	(0.0529)
R^2	0.256	0.233	0.144	0.121	0.097	0.079
Control mean	0.5	0.36	0.18	0.09	0.25	0.19
Panel B: Placement outcomes weighted by individual preferences						
Treatment	0.264**	0.175	0.237**	0.244**		
	(0.112)	(0.111)	(0.118)	(0.122)		
N	293	293	293	293		
R^2	0.204	0.182	0.155	0.113		
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Batch FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1)-(3) report placement outcomes and column (4) reports whether the trainee is in the same job three months later. Panel A reports the regression outcomes with unweighted placement outcomes whereas Panel B weights the outcomes by trainee preferences and normalizes them to have mean 0 and standard deviation 1 for the control group. Individual controls used are the number of interviews, age, gender, years of education and dummies for whether the trainee is a student or not and whether from a SC/ST/OBC caste category. Robust standard errors are reported in parentheses. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

B Treatment effect with batch-level and individual-level randomization

As discussed in the paper, a limitation of our current experimental design (henceforth experiment #1), where we randomize individuals within a batch into treatment and control, is that we do not observe a counterfactual allocation of interviews in the absence of our treatment. As discussed in section 6, this implies that we cannot directly examine whether our intervention increased aggregate welfare once we incorporate spillovers in the control group. An alternate experimental design (henceforth experiment #2) could be to allocate treatment across batches (instead of within them). This would therefore ensure we have a control group of batches to empirically observe the allocation of interviews in the absence of our intervention. In this section, we show however, that we would require double the number of batches under experiment 2 as compared to 1.

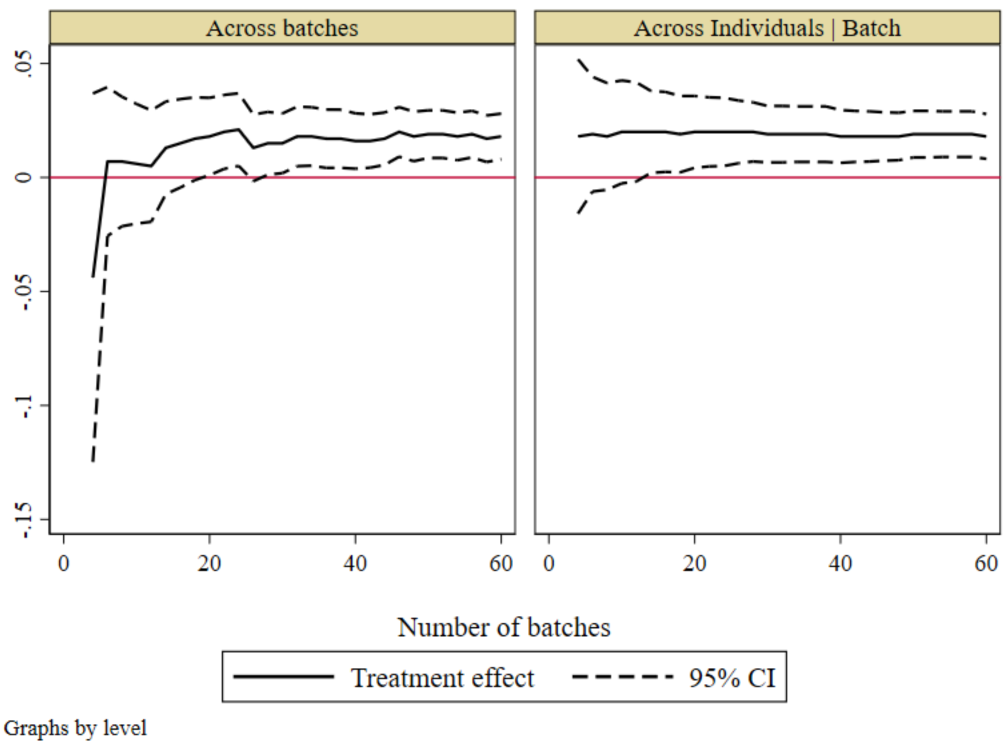
We implement the following steps to estimate the treatment effect under both experimental designs, where the treatment effect is the differential probability between the treatment and control trainees of getting an interview for at least one of their four most-preferred jobs.

1. We sample (with replacement) $B = \{2, \ldots 60\}$ batches from the empirical distribution of batches (and trainees within them). Under design #1, within each batch, we randomly allocate half the trainees to treatment, while the other half are control. Under design #2, half the batches (and hence all trainees within them) are allocated to treatment, while the other half are control.

2. We then allocate the interviews (which we observe empirically) using the allocation mechanism discussed in section 6.1. Specifically, as discussed previously, we assume that the manager's information set is the "no information" case for the control trainees/batches and the "hybrid information" case for the treatment trainees/batches. Simulating the algorithm 25,0000 times therefore allows us to measure the probability that a trainee i gets allocated a at least one job in her four most-preferred jobs.

For a sample of B batches therefore, we can estimate the treatment effect under both experimental designs. This is what we report in figure B1. The horizontal axis reports the number of batches in our sample and the vertical axis reports the treatment effect coefficient. As can be seen in the figure, detecting the treatment effect becomes more precise with more batches. However, the first experimental design requires almost half the number of batches (less than 20) to detect a treatment effect at conventional levels, as compared to the second experimental design (around 40).

Figure B1: Treatment effects under batch-level and individual-level randomization



Notes: The horizontal axis reports the total number of batches, while the treatment coefficient (with its 95% confidence interval) are reported on the vertical axis. The first graph reports the treatment effect when the treatment is allocated across batches, while the second graph reports the treatment effect when the treatment is allocated across individuals within a batch.