PEDL Research Papers

This research was partly or entirely supported by funding from the research initiative Private Enterprise Development in Low-Income Countries (PEDL), a Department for International Development funded programme run by the Centre for Economic Policy Research (CEPR).

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

Abhijit Banerjee † Gaurav Chiplunkar ‡

MIT Yale University

August 2018

Abstract

This paper provides evidence of substantial matching frictions in the labor market in India. In particular, placement officers in vocational training institutes have very little information about the job preferences of candidates they are trying to place in jobs. In the first part of this study, we adopt several alternative methods to elicit genuine preferences of candidates over different types of jobs and show that placement officers have poor knowledge of these preferences. In the second part, we provide placement officers with this information and examine its impact on placement outcomes and employment. We find that placement officers come closer to efficiently matching candidates to job interviews. Based on estimating a structural model of job preferences, we argue that there are net welfare gains because of better matching, not just redistribution within the group of potential employees. Furthermore, this leads to substantial improvement in job choices made by the candidates and subsequent employment outcomes for three to six months after initial placement.

*We thank Esther Duflo, Glenn Ellison, Clement Imbert, Parag Pathak and Debraj Ray 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.

†banerjee@mit.edu

‡gaurav.chiplunkar@yale.edu

1 Introduction

An important but under-emphasized fact about the Indian economy is captured in figure 1. It uses data from the 66th round of the National Sample Survey (2009), 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, 20.2% of the more educated young men are not employed while the same number among the less educated men is 1.8%. The difference is significant at the 0.001 level. Figure A1 in the appendix shows the breakup of the non-employment rate between education and seeking employment. About half of those who are not working at age 25 (51.09% to be exact), which is 10.6% of the population of that cohort, claim to be available for work (though perhaps not all of them are actively looking for a job). Moreover among the rest of the non-working population, 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 those who have more than 10 years of education and those who have less than 8 - both have non-employment rates of around 0.3% with a p-value of 0.76. While these could be cohort specific differences, we see very similar patterns in figure A2 in the appendix, which reports on the two previous rounds of the survey that collected the same data (the 43rd round and 55th rounds from 1987 and 1999 respectively).

This suggests two possible stories about why there are so many job-seekers. First that the search mechanism is inefficient and it takes a long time for people to find the particular job they want. Second, people start by aiming high in the job market and slowly adjust their expectations based on their experience. This could be entirely rational- if certain jobs have lots of rents, it may make sense to focus on getting one of those jobs rather than settling for a bad job immediately

[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.

after college- but only if the probability of getting one of those jobs is high enough. We will return to this question in the concluding section.

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 tax distortions make it possible that the individual job seeker searches too little, which would justify incentivizing 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. They report that these strategies are on average successful in raising labor market outcomes for those who are exposed to them, though it is not clear how much of this is displacement rather than net gain for the entire labor market.[2] However, almost none of these interventions are carried out outside the OECD. Card et al. (2010) report that only 2% of the 132 impact estimates they have for non-OECD countries are for search assistance programs, despite the numbers cited above for India and the even higher non-employment rates among the young in many developing countries including South Africa and the MENA region.

This paper reports on a randomized trial of an intervention during the job placement process of a large Indian vocational training firm. It starts by providing detailed evidence for an important source of mismatch: placement managers (who are responsible for matching job seekers to interviews) 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 clearly need to reliably know the preferences of each job seeker– otherwise what we may believe to be mismatch could reflect the fact that the placement manager knows more about the preferences than we do. Unfortunately, getting people to reliably reveal their preferences is not easy, especially when the preferences are multi-dimensional, so that standard BDM mechanism cannot be used. To elicit preferences of job seekers over a set of job

[2] Crépon et al. (2013) find large displacement effects from job placement assistance.

characteristics, potential employees (who are currently trainees in a job-skills training center) are asked to make choices by ranking some real job options and some invented ones mixed in with them. The jobs that they rank are carefully chosen to exploit variation along different job characteristics. Trainees are told and it is in fact true that they are more likely to get interviews at jobs that they rank higher, which is where our partnership with the training firm helps us. To test whether these choices reflect their true underlying preferences, for half the trainees we emphasize that the probability they will get one of these interviews is high (which is once again 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 report strategically to maximize their chance of getting a job. Finally, we make them list the attributes of the jobs they like: it turns out that the preferences revealed by just asking them this are very consistent with their preferences elicited through the more elaborate job choice exercise.

Having thus confirmed that we know what their true preferences are, we ask the manager for a training center, who is responsible for the placements of these trainees, to predict the preferences of each trainee over the same set of jobs used for the trainees' preference elicitation- specifically we ask her 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 choice correlates with the trainee's preferences, this allows us to examine the extent to which the person in charge of placement knows the preferences of the person they are placing.

The results are consistent with the center managers having lots of information about some trainees but very little information about others. 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 the cases. On average, the job picked as the best job for a particular trainee by the manager was ranked at 7.2 by the candidate himself on a scale of 1 to 11 (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 would have been 11. So it appears to be the case that the manager does do slightly better than random choice, but is far from knowing their trainee's preferences.

Having documented the lack of knowledge of trainees' preferences by their managers, the second part of the study, starting in section 5, is a randomized control trial. We experimentally vary the information that the placement managers have about the preferences of the trainees and show that this substantially improves the matching between trainees and the interviews they get. However, this is a very partial equilibrium view. Though the trainees in the treatment group benefit from the information being provided to the manager, it does not mean that the overall matching has become more efficient or desirable in any way– this is what we explore in section 6. We make assumptions about the 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 about their preferences, (iii) a *hybrid information* case, where she knows what we tell her about the treatment group but what she tells us about the control group. Under these alternative assumptions, we ask whether a stable matching algorithm can predict what 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 that the hybrid information case perhaps fits the data the best. In other words, the manager in this sense, comes close to achieving efficiency subject to her information constraints. With further assumptions, we are able to quantify the utility gains from our experiment for both the control and treatment trainees. We find that though the ex-ante expected utility decreases by around 6% for the control trainees, it increases by over 15% for the treatment trainees, indicating that *after* taking into account the potential reallocation of interviews across trainees, there is a net gain from our intervention. The final section of the paper asks whether the success in altering the matching has labor market consequences. The answer seems to be yes: the intervention does seem to have large and significant effects on trainees getting job offers, accepting offers for jobs that they prefer and staying employed in them for at least three to six months after the initial placement.

This paper makes various contributions to the existing literature: in terms of studying the delivery of active labor market policies, this is obviously related to the set of recent papers comparing public and private job counseling services in OECD countries. Both Krug and Stephan (2013) in

Germany and Behaghel et al. (2014) in France show evidence from randomized controlled trials to the effect that public services work better than outsourced private services, while Laun and Thoursie (2014) find no difference between the two.[3] Behaghel et al. (2014) argue that this reflects better incentives and higher competence in the public sector. We cannot say whether the failure that we detect is a matter of competence or incentives because we focus one very specific aspect of the placement manager's job, but we have much more precise evidence of where they are failing and therefore can argue that it is extremely inexpensive to fix.

In terms of identifying a very specific (but very different) distortion in the job matching process, this paper is one of the first to experimentally show the importance of information frictions (and its labor market consequences) where labor market intermediaries play an important role in matching firms and job seekers. In this regard, the paper contributes to a recent literature that has focused on testing interventions to reduce search and information frictions between workers and firms like for example, Dammert et al. (2015), who provide information on vacancies to job seekers in Peru; Beam (2016); Abebe et al. (2017), who test the impact of job fairs; Franklin et al. (2015), who provide 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.

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 efficiency of the matching process and its welfare 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.

2 Context and background data

[3] Bennmarker et al. (2013) find that outsourced services work slightly better, but in their case the intensity is higher in the private case.

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 the firms indicate facing shortage of labor and 89% 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 finds 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

[4] FICCI Survey on Labor/Skill Shortage for Industry, October 2011.

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 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[6]. As can be seen in column (1), our study sample is young

[5] http://theskillsacademy.in

[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.

(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 them one by one and then put them together to check if the two procedures give similar results.

3.1 Hypothetical choices

Job aspirations

In a survey implemented in the first week of the training program, we asked trainees about their aspirations with regard to employment after the training program. We focused specifically on four aspects of a job that from other accounts, are important for a trainee: employment sector, location, salary and whether there is a provident fund (PF), which is a mandatory savings scheme where the firm is required to match the employees contribution.[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. 78% 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

[7] 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" company,

[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.

preference for a job with provident fund. Panel C reports the location preferences, which is broken down based on the residence of the trainee. For the 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. 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.

Job priorities

In the same survey as above, trainees were asked directly about their preferences over different job characteristics by asking them to distribute a hundred points across various job characteristics. Each trainee was presented with six job characteristics[9] and was asked to distribute a hundred points across them. 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 genders is with respect to location, which is more important, perhaps not surprisingly in the Indian context, for women than for men.

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.

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

Incentivized elicitation of preferences

To begin, we generated a list of jobs by varying the job characteristics that trainees reported as important in the hypothetical activity above, namely: 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 the 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]. The variation in the job characteristics is summarized in figure 2. 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 that were available to these trainees. To ensure this, within every employment sector that trainees were 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. 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. Figure 3 provides an example of one such job list that was presented to the trainees and figure 4 is an example of one particular job (a job for a receptionist in Lucknow that pays Rs. 6,000 and where no provident fund is provided).

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 true preferences rather than choosing strategically to maximize their chance of getting something, since our objective was to get people to jobs that they would genuinely want and therefore retain. This suggested making the stakes less salient. In the end, we decided to go for

[10] 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.

the two extremes, with the view that if they yielded more or less the same result, we could be reasonably confident that we have what we need and if they differ we would try to combine in some way. More specifically, a randomly chosen half of trainees within every training batch 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 get. The other half of the trainees in the same training batch were told that there was a very high likelihood that their job rankings would determine the interviews they get. In both cases, because of our partnership with Skills Academy, the description was factually correct.

We now come to the results of this activity. One primary challenge we faced in implementing this exercise 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 (which is 63% of the sample). Table A1 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 A2 reports a standard balance check on the observable characteristics of trainees assigned to low and high salience groups. We find no statistical difference on observable characteristics between these two groups. Finally, columns (5)-(7) of table 4 present the results on ranks given to the *same* set of jobs by trainees in the two groups. As reported in the table, there seems to be no difference in the average rank given to a job by trainees in the two groups–the differences are both small in magnitude and nowhere near statistically significant. Going forward, we will therefore assume that these job rankings reflect the true underlying preferences of trainees over jobs.

We now examine the heterogeneity of preferences across the 11 jobs in table 4. For each of the 11 jobs that trainees ranked, we calculate what fraction of trainees that placed the job in the bottom three jobs (column (2)), ranked the job in the middle i.e. between 4-8 (in column (3)) and finally, ranked the job among the top three jobs (column (4)). We see that there is a substantial heterogeneity in preferences across trainees. For example, more than 30% put jobs 2, 3, 4, 8 and 9 among their bottom three jobs but another 18% or more 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 the jobs based on preferences.

Compensating differentials

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

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

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)[11] 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 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 lastly, for jobs in the rest of the country, we deflate the salary 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

[11] 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

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 the results 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 significant for men. However it is notable that while men seem to be almost indifferent between desk jobs and active jobs (e.g. delivery) 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 significant at conventional levels. Consistent with the stronger preference for staying local among women, the in-state premium is 54 percent for men and 136 percent for men relative to staying in their home district. This is what we would have expected given the social context of North India.

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 hence 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 and that an exercise of hypothetical elicitation of job preferences can be indicative of the actual preferences.

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 officers 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. For measuring trainees' preferences, we use two metrics: (i) the ranking of jobs as described in section 3.2 and (ii) hypothetical preferences generated from their stated job priorities as described in section 3.3.

Using the manager and trainee preferences, we construct four measures of "how well" a manager knows her trainee's preferences (described below). As a benchmark, we can compare each of our measures to two hypothetical scenarios: one where the manager responded with a random list of jobs and one where the manager has perfect knowledge of the preferences of the trainee and responds based on that. The results for this activity are reported in figures 6,7 and table 7. We now discuss the four measures in detail below:

1. <u>Measure #1:</u> 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 if we use the job ranking and 5.44 if we use the hypothetical preferences. Using job ranking does significantly better than the random process whereas using hypothetical preferences does no worse than the random process. Both preferences do significantly worse than the case where preferences were known perfectly.

2. Measure #2: 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 if we use job rankings, which is significantly better than the random process, but far worse than the perfect information case. Using the hypothetical preferences, the average trainee rank is 4.78, which is significantly worse than even the random process.

3. Measure #3: 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 by using the job rankings, which is significantly better than the random mechanism, but far below the perfect information case. The average is 7.26 by using the hypothetical preferences, which is significantly worse than even the random process.

4. Measure #4: 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 and the average is 0.17 if we use the hypothetical preferences. Both correlations do significantly 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 centre 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. She does do slightly better than choosing completely at random, but is nowhere near perfect information. Furthermore, as shown in figure 6, even across trainees, there seems to be a considerable amount of variation in the knowledge of manager. For example if we use the job ranking, in 20.5 percent 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 percent 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

The intervention was as follows: trainees in each cohort 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. For the second group (henceforth the Control group), no trainee preferences were shared with the manager. In figure 5, we give an example with information about two such profiles that were presented to the center manager. Table A3 checks for balance across trainee characteristics between the control and treatment groups to test for the randomization. They are balanced on observable characteristics. 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 this intervention is to reduce the asymmetry of information on trainees' preferences over the set of firms.

5.2 The impact on the number and type 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:

$$y_i = \alpha_b + \beta T_i + \gamma X_i + \varepsilon_i \quad (2)$$

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. 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; α_b is a batch/cohort fixed effect (since students were trained in batches of 20 or so). y_i in column (1) in table 8 is a dummy variable that takes the value 1 if the trainee received at least one interview and conditional on getting at least one interview, in column (3), the number of interviews. Column (2) has the number of interviews, which is equal to 0 if the trainee received no interviews. As reported in columns (1)-(3), there are no differential effects of the treatment on the number of interviews.

We then examine whether the type of interviews, as measured by the characteristics of the job were different between the treatment and control. We run the same specification as in (2) where the dependent variable y_i is now a dummy variable for salary, location and PF categories respectively. The results are reported in columns (4)-(6) of table 8. Again, there are no differential effects between the treatment and control trainees. [12]

5.3 Quality of interviews: data challenges

Given that there is no effect on the number or types 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 cannot 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,

[12] The Cohen's d effect sizes calculated from the regression coefficients are 6.94%, -8% and -6% respectively. The experiment has power to be able to detect effects of 25%, 21.5% and 19.2%.

location and provident fund.

The bigger challenge was that if we take the complete set of combinations along the three dimensions (salary, location and provident fund) we would have 18 potential jobs. However, 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 141 interviews (65%) with those in the job ranking list. However, for the remaining jobs, we do not have a match (and hence we do not know the preferences of the trainee). Going forward, we only consider the interviews where we know the trainee preferences.[13] The last row of table A3 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 information for the top four preferred jobs of the trainee to the manager, we examine the impact of this intervention on two outcome variables: (i) a dummy variable for whether the interview was in the four most preferred jobs of the trainee and (ii) the (normalized) rank[14] for the interview as reported by the trainee in the ranking exercise. For both the outcome variables, we then estimate the following OLS regression at the person-interview level (hence conditional on getting an interview and within the set of matched interviews):

$$y_{ij} = \alpha_b + \beta T_i + \gamma X_i + \varepsilon_{ij} \tag{3}$$

The results are reported in columns (1) and (3) of table 9. Column (1) tells us that the provision of information to managers has positive (but not significant effect) on the matching trainees to jobs that were more preferred. However, when we re-estimate it using a logit specification for the

[13] In an alternate exercise, we use LASSO to predict preferences for all interviews and hence redo our analysis using all interviews instead of just the ones where we have an exact match. Qualitatively, the results are the same but since the estimated preferences are very noisy, we lose statistical precision in our analysis.

[14] We normalize the rank for the interview to have mean 0 and standard deviation 1 so that the regression coefficient can be interpreted in terms of standard deviations.

limited dependent variable as reported in column (2), the provision of information on preferences increases the odds of getting an interview by 52.9%, which is large and significant at the 10% level. Furthermore, as reported in column (3), trainees in the treatment group interviewed for jobs that were on average 0.32 standard deviations more preferred than those in the control group and this effect is significant.

The fact that we could not rank all the jobs means that we are probably underestimating the treatment effect in columns 1 and 2. In many of the cases where the manager had given the trainee interviews that best approximated what the trainee wanted and would in fact be in his top 4, we would not count it because it was not an exact match.[15]

6 Welfare consequences

One problem with interpreting these results as evidence of the success of our intervention is that they may have actually made things worse on average when one includes the control group. This is because we gave the managers information about the preferences of roughly half the people they had to assign interviews to while saying nothing about the others. This can easily lead the manager to move to an allocation which is worse on average and from one that is in the core to one which is not.

For example let there be 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 $\{a \rightarrow: 3; b \rightarrow 1; c \rightarrow 2\}$. b gets what manager's best information says should be her top choice. Now suppose the manager is now told 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 $\{a \rightarrow 1; b \rightarrow 3; c \rightarrow 2\}$. 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 just went down by one, while the number of people with the top preference is still one.

[15] 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.

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 in aggregate welfare went up or down. To make progress and get at the welfare consequences of the experiment, we need to be able to predict which interviews the individuals in the treatment and control groups would get absent the intervention. For this we need to come up with a model of the manager's decision rule based on the observed allocation of interviews in treatment and control. This is what we do in 6.1. Next, assuming that this rule is a reasonable approximation to how the manager actually decides, we can generate the counterfactual allocation for individuals in treatment and control absent the intervention. Finally, in section 6.2, we impose a functional form on the utility function to compute the net utility gain from the experiment for trainees in the control and treatment groups.

6.1 The manager's decision rule

Information sets of the manager

We first begin by restricting the information sets about trainee preferences that the manager could have when she allocates jobs to trainees. First, we can consider a *complete information* case, where the manager knows the preferences revealed in the job ranking exercise (from section 3.2). Second, we can go to the other extreme *no information* case and base the matching exercise on what the manager *thinks* are trainees preferences (from section 4),[16]. This is a reasonable benchmark for what a manager would do if she cannot process the information we gave her about the preferences. Finally, we can construct a *hybrid information* set 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.

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

Job allocation rule

To assign a decision-rule to the manager, we assume that the manager chooses allocations of interviews that would be in the core under each of the hypothesized information sets above. We can then compare the predicted allocations under each hypothesized information set with the actual allocation in order to choose the information set most likely to be the one that the manager uses.

Before we proceed, we would need to clarify several things related to the matching algorithm: first, we assume that the manager has no preferences over which trainee should get which interview.[17] Second, we implement the following algorithm to identify allocations that are in the core: trainees in a batch are arranged in a random order; we then allow the first trainee to pick a job from the set of available jobs. Then the next one picks from the remaining jobs and so on. Third, for almost all batches there are more trainees than interviews–so any matching algorithm would have multiple matching allocations that are in the core (and hence stable). To take this into account, we run the algorithm 25,000 times, each time ordering the trainees randomly within each batch to simulate the set of stable allocations and thus calculate the probability that a trainee i is matched to an interview for job j. However, a "jobs" in our setting is purely defined by the salary, location and whether or not there is a provident fund. A variation in any other dimension (work timings for example) is not captured. As a result, we observe some people getting multiple interviews for the same "job". We then sum the probabilities across all the jobs classified as the "same" job to calculate the probability that a trainee i is matched to any interview for job j. Fourth, we observe some individuals being matched to multiple interviews for the same job. So unless we make further assumptions on how individuals can trade "bundles" of interviews, we cannot perfectly compare the theoretical and empirical outcomes since in the simulated outcomes every individual gets only one job. For our main results, we therefore only consider the batches where less than 20% of the trainees get more than one interview. As shown in figure 8, even with this restriction, we are able to examine allocations in 19 out of the 21 batches.[18]

[17] 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.

[18] We do a robustness check where we include all batches and can show that our qualitative results do not change.

Results

With these caveats in mind, under *each* of the three information sets of the manager, the matching algorithm then generates a probability that an individual i is matched with an interview for job j, which we denote by p_{ij}. The goal of this exercise is to then compare p_{ij} to the actual allocation of interviews. In the data, we create a dummy variable (D_{ij}) that takes a value 1 if a trainee i actually gets an interview j and 0 otherwise. Pooling all the interviews and trainees, we can therefore calculate $E(D_{ij}|p_{ij})$, which is the expected probability of *actually getting* an interview conditional on the theoretical probability that a trainee *should* get one according to the matching algorithm. Figure 9 plots this relationship. If managers know trainee preferences perfectly, this should coincide with the 45 degree line. However, as can be seen in the first graph of figure 9, for low values of p_{ij}, the empirical allocation is not very informative about the manager's information set. On the other hand, there is a stark difference in the allocation efficiency for higher values of p_{ij} with the hybrid coming much closer to fit the data. It is important to note however, as shown in the second graph of figure 9, most trainees have relatively low values of p_{ij}, which is not surprising given the scarcity of jobs. It is jobs that very few people want where the manager information seems to make the biggest difference. This is intuitive, since for these jobs being able to identify the small number of people who really want them creates a potential for a large welfare gain.

Discussion

The above exercise tells us two useful things. First, assuming that the manager knows trainee preferences does a bad job at predicting how managers allocate interviews, which indicates there is a gap between the manager and the trainee in the knowledge of trainee preferences. Second, the hybrid information set seems to do the best, suggesting that the manager does use the information we provide her with.

6.2 Welfare implications

Our intervention could simply result in a reallocation of interviews, making the treatment group better off on one hand, it could make the control group worse off on the other. We now turn to measuring the welfare gains from our intervention *after* taking into account the reallocation of interviews. We begin by assuming utility (U_{ij}) that an individual i (with characteristics X_i) gets from job j (with characteristics Z_j) take the following form:

$$\begin{aligned} U_{ij} &= \beta_i Z_j + \varepsilon_{ij} \qquad (4) \\ \text{where: } \beta_i &= \bar{\beta} + \gamma X_i \\ \varepsilon_{ij} &\sim \text{i.i.d Type 1 EV} \end{aligned}$$

We can then use a rank-ordered logit framework as in Beggs et al. (1981) and Allison and Christakis (1994) to consistently estimate $\bar{\beta}$ and γ by a maximum likelihood estimation. We take Z_j to be discrete categorical values for salary, location and provident fund. For the vector of individual characteristics X_i, we include dummy variables for female, college and urban along with the age of the individual. Table 10 reports the estimated values of β_i. By using these estimated values, we can then predict the deterministic part of utility $\hat{\beta}_i Z_j$. Now to calculate utility, the rank ordering of jobs for an individual gives us information that can be used to put bounds on the unobservables ε_{ij} that affect utility. So for a simulation s, we calculate utility $U_{ij}^{(s)}$ as follows: (i) for each individual i, we start with her lowest ranked job r_{i1} and draw $\varepsilon_{i1}^{(s)}$ from a Type 1 EV distribution. For all other jobs $j = \{r_{i2,} r_{i3} \dots r_{iJ}\}$ since $r_{ij} \geq r_{ij-1}$, from (4), it must be the case that $U_{ij} \geq U_{ij-1}$, which implies that $\varepsilon_{ij}^{(s)} \geq \varepsilon_{ij-1}^{(s)} + \hat{\beta}_i(Z_{j-1} - Z_j) \equiv \underline{\varepsilon}_{ij-1}^{(s)}$. Therefore, given a draw of $\varepsilon_{ij-1}^{(s)}$, we draw $\varepsilon_{ij}^{(s)}$ from the Type 1 EV distribution truncated below by $\underline{\varepsilon}_{ij-1}^{(s)}$. Utility $U_{ij}^{(s)}$ is then simply $\hat{\beta}_i Z_j + \varepsilon_{ij}^{(s)}$. Lastly, we average across 1000 simulations and define utility of an individual i for job j to be: $U_{ij} = \frac{1}{S}\sum_s U_{ij}^{(s)}$.

We can now measure the (ex-ante) expected welfare gains from our intervention across trainees in the treatment and control. First, as discussed in section 6.1, since the hybrid information case approximates the actual allocations well, we can use p_{ij} estimated under the hybrid information case and U_{ij} from above to calculate the expected utility for an individual i, defined as $V_i^H = \sum_j p_{ij}^H U_{ij}$. To measure the "gains" from our intervention, it is reasonable to assume that allocations generated in the no information case (discussed in section 6.1) are a good approximation for allocations in the absence of our intervention. Similar to above, we can then use p_{ij} generated under the no information case to calculate V_i^N, which would be the expected utility for an individual in the absence of our intervention.

We measure the gains from our intervention by examining the expected utility for treatment and control trainees under our intervention (V_i^H) as compared to without it (V_i^N). This is shown in figure 10. Both graphs plot V_i^N on the horizontal axis and V_i^H on the vertical axis and the 45 degree line in red. Each dot represents $\{V_i^N, V_i^H\}$ for a trainee. The first graph on the left plots the welfare gains for trainees in the control group while the second graph plots gains for trainees in the treatment group. As can be seen from the graph, on average, expected utility of trainees in the control group is lower with the intervention than in the absence of it (dots are below the 45 degree line). In the treatment group, there is a greater variation where some trainees gain from the intervention while other lose out. However, on average, welfare is 6.28% lower for control trainees and 15.47% higher for treatment trainees with our intervention than in the absence of it. This indicates that though our intervention results in a reallocation of jobs (both from the control to the treatment as well as within treatment), the expected utility gains in the treatment group on average outweigh the loss in the control group, under the admittedly strong assumptions we make.

7 Impact on job acceptance and employment

The above analysis provides us with evidence that the intervention of giving managers the job preferences of trainees did have an impact on improving the efficiency of the matching process as well as had welfare gains on average assuming that the interview was the final outcome. In this section, we explore whether this further resulted in improving actual job outcomes. Since we observe (a) the reported preference by a trainee for a job; (b) various placement outcomes[19] for every trainee-job pair, we can use them to examine the impact of our intervention on job quality and subsequent job retention.

[19] For the jobs that the trainee had ranked, but got no interview, we set all outcome variables to zero.

7.1 Treatment effects

With the 141 interviews that we can exactly match for a trainee i and job j, we consider three outcomes related to interviews and offers– (i) the number of interviews; (ii) number of offers; (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. We first begin by aggregating all outcomes to create an individual specific outcome variable across all jobs (denoted by y_i) and estimating the following regression specification:

$$y_i = \alpha_b + \beta T_i + \gamma X_i + \varepsilon_i \tag{5}$$

where: y_i are a set of job choice and placement outcomes, T_i is a dummy variable that takes the value 1 if the trainee was in the treatment group and 0 for the control group and X_i are the set of individual controls used in previous regressions. From the results reported in table 11, we do not see any effect of the treatment on any of these outcomes.

This result is consistent with our previous discussion since the intervention really improves the *quality* of interviews for the treatment group and not the *number* of interviews itself, though perhaps we would have expected an improvement in job retention (see later). Hence, we now examine the differential impact of the treatment based on how much a trainee prefers being matched to that interview. We can do this in multiple ways: first, we consider outcomes at the trainee-job level and examine the heterogeneity of the treatment based on the trainee's preference for that job. Second, we create an index of *placement quality* by weighting all outcome variables at the trainee-job level by the trainee's preference for that job and aggregating it across jobs to create an index of aggregate placement quality for each trainee. We discuss both these approaches in the subsequent sections below.

7.2 Heterogeneity of treatment by job preferences

Taking all outcome variables at the trainee-job level, we estimate the following regression specification:

$$y_{ij} = \alpha_i + \beta P_{ij} + \delta T_i \times P_{ij} + \varepsilon_{ij} \tag{6}$$

where: y_{ij} are now the set of outcomes at the trainee-job level and P_{ij} is the (normalized) job rank as reported by trainee i for job j in the job ranking exercise. The trainee fixed effect (α_i) controls for all observed and unobserved trainee characteristics. It also absorbs the direct effect of the treatment and therefore the coefficients are estimated using the variation in preferences across jobs within a trainee.

Table 12 reports the results for the above regression specification.[20] As a benchmark to compare the coefficients, we use the mean of the outcome variable in the control group. First, as reported in column (1), trainees in the treatment group got 0.0252 additional interviews that were ranked one standard deviation higher as compared to trainees in the control group. Considering a benchmark in the control group where the average quality weighted number of interviews was 0.051 (across all trainee-job pairs), this translates into a 49.3% increase in the quality weighted number of interviews. This result concurs with our previous analysis in section 5.4 that the intervention matched treatment trainees to interviews that were more preferred.

Second, we examine whether this had an impact on the number of job offers received by trainees. As reported in column (2), relative to the control group, trainees in the treatment group got 0.014 more offers for jobs that were ranked one standard deviation higher. This is a 44.1% increase as compared to the benchmark average of 0.03 offers across trainee-job pairs in the control group.

Third, we consider the likelihood of a trainee accepting a job offer for jobs she prefers more. The outcome variable is therefore a dummy that takes the value 1 if the trainee accepted a job offer and 0 in all other cases (even if the trainee got no interviews/offers at all). As reported in column (3), trainees in the treatment group were 1 percentage point more likely to accept offers for

[20] We redo our analysis by taking all the interviews and preferences predicted using LASSO and find that qualitatively, nothing changes, though the results are statistically noisy.

jobs with one standard deviation higher rank relative to the control group, which is a 92% increase as compared to the benchmark average acceptance rate (unconditional on getting an offer) in the control group.

In columns (4)-(7) of table 12, we examine the impact of our treatment on job retention and employment outcomes. Given that we had two rounds of follow up surveys, three and six months after the completion of the training program, we can examine the persistence in the impact of our intervention over six months.[21] In columns (4) and (5), the outcome variable is a dummy that takes the value 1 if the trainee was employed in the *same* job three months and six months later respectively and 0 otherwise. Columns (6) and (7) construct a similar dummy variable to examine if the trainee was employed in *any* job. As reported in column (4), trainees in the treatment group were 0.76 percentage points more likely to be employed in jobs that they ranked one standard deviation higher as compared to the control. This translates into a 125% increase in the retention of a trainee in a job as compared to the average in the control group. Column (5) looks at job retention after six months where we find a large effect in job retention, but the result is not statistically significant. Lastly, as reported in columns (6) and (7), trainees in the treatment group are 13.5% and 36.4% more likely to be employed in any job (not quality weighted) three and six months later respectively, though this difference is not statistically significant (with a p-value of 0.51 and 0.18 respectively). While not significant at conventional levels the 6 month effect is large. Combined with the absence of a job retention effect, this suggests that once people find a job they like, they either perform well and therefore find a better job, or are enthused to look for a better job.

7.3 Impact using a placement quality index

A second way of examining the impact of the intervention is to create an index of *placement quality* for every individual. We do this as follows: for each trainee i and job j, we weight the outcome variable y_{ij} with trainee i's (normalized) rank for that job j (P_{ij}) and sum it across all jobs to create an index for that outcome for trainee i (denoted by Q_i^y). Therefore:

[21] The number of observations in column (5) is lesser because we were only able to survey 90% of our trainees after six months.

$$Q_i^y = \sum_j P_{ij} y_{ij}$$

We can then examine the impact of our intervention on improving this index (by using Q_i^y as the dependent variable). Therefore, for all placement outcomes discussed above, we estimate the following specification:

$$Q_i^y = \alpha_b + \beta T_i + \gamma X_i + \varepsilon_{ij} \tag{7}$$

where Q_i^y is the placement quality index for outcome y as defined above and the other variables are the same as used in previous regressions. The results are reported in table 13. The results concur with our previous analysis– as compared to the control, trainees in the treatment group got better quality interviews and offers, were more likely to accept them and retain them in the short term (three months).

8 Conclusion

This paper identifies an important potential source of mismatch in the Indian labor market – that intermediaries (center 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 agenda of understanding search costs and mismatch in the labor market and ways to reduce them. While others have emphasized externalities (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 the important work of Bloom et al. (2013) in understanding the inefficient management practices in India. Understanding why managers do not use this information or at least try to gather it is the next step on our research agenda.

Going beyond the specific issue of the 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 in the results 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.

Beyond that it may be important to start a culture of unpaid internships in firms for high school 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.

Allison, Paul D and Nicholas A Christakis, "Logit models for sets of ranked items," *Sociological methodology*, 1994, pp. 199–228.

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.

Beggs, Steven, Scott Cardell, and Jerry Hausman, "Assessing the potential demand for electric cars," *Journal of econometrics*, 1981, *17* (1), 1–19.

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.

Bennmarker, Helge, Erik Grönqvist, and Björn Öckert, "Effects of contracting out employment services: Evidence from a randomized experiment," *Journal of public economics*, 2013, *98*, 68–84.

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

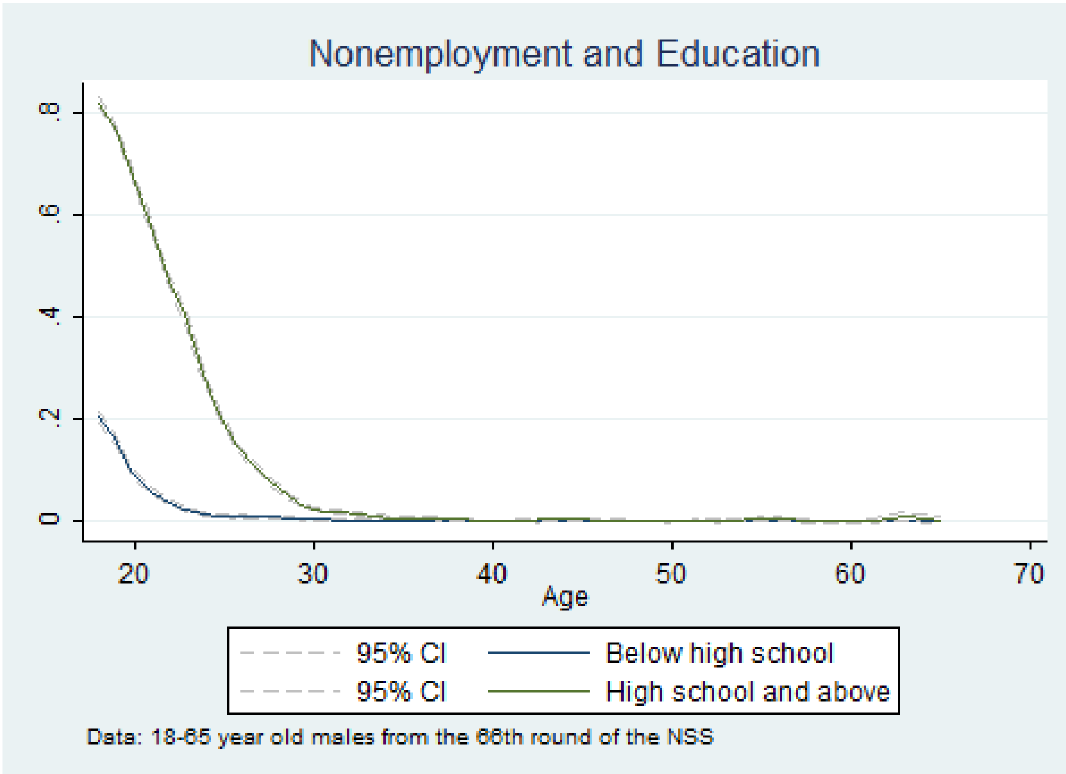

Figure 2: Variation in job characteristics

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

Figure 3: Job list for ranking (Example)

Name:

Gender:

Centre:

Trade:

Group:

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: ________________

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: ______________

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: ______________

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: ______________

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: ______________

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

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

Figure 4: Example of a job

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

Figure 5: Example of preferences 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)

Figure 6: Manager's knowledge of trainee job rankings

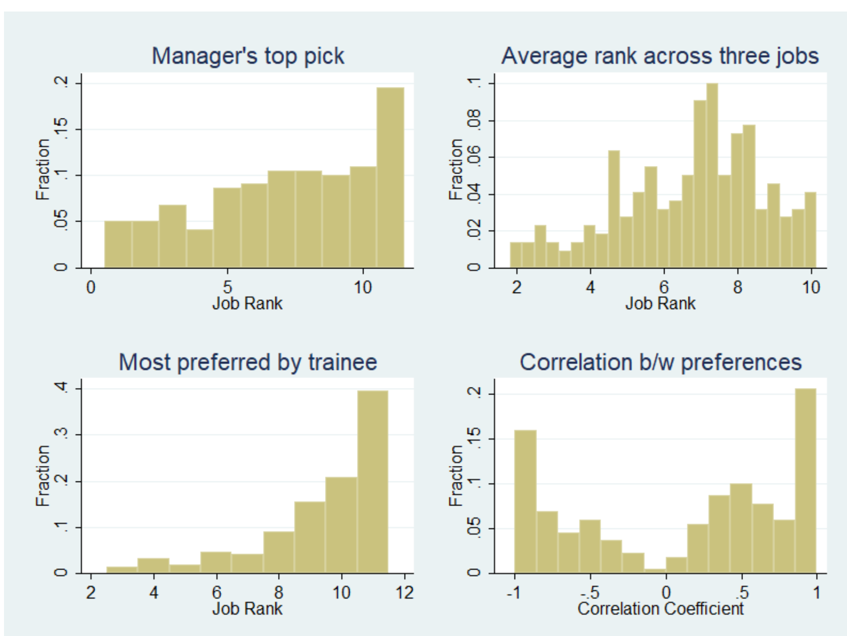

Figure 7: Manager's knowledge of hypothetical trainee preferences

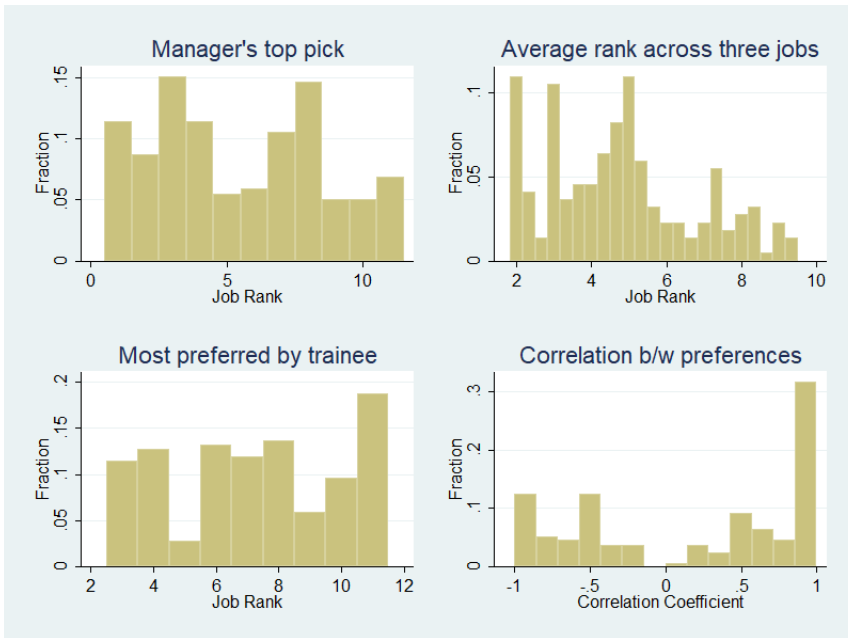

Figure 8: Distribution of trainees and interviews

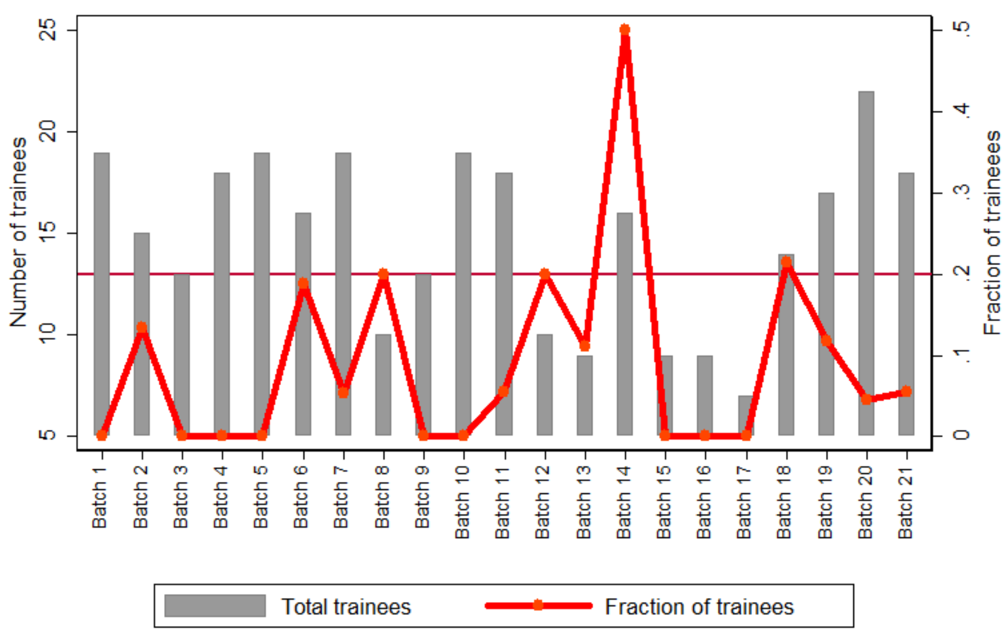

Figure 9: Stable matches and actual outcomes

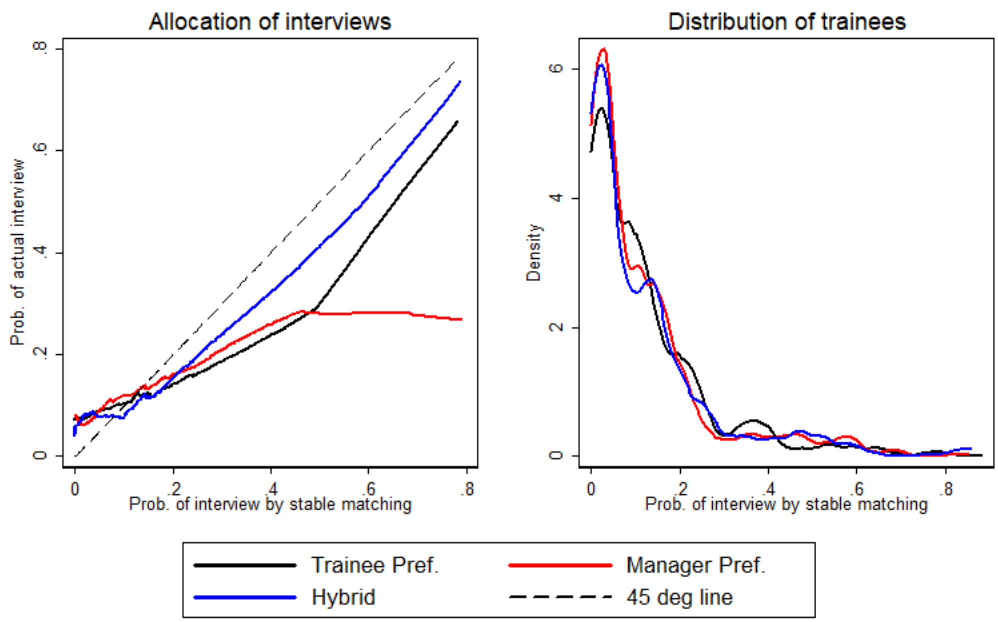

Figure 10: Utility gains across control and treatment

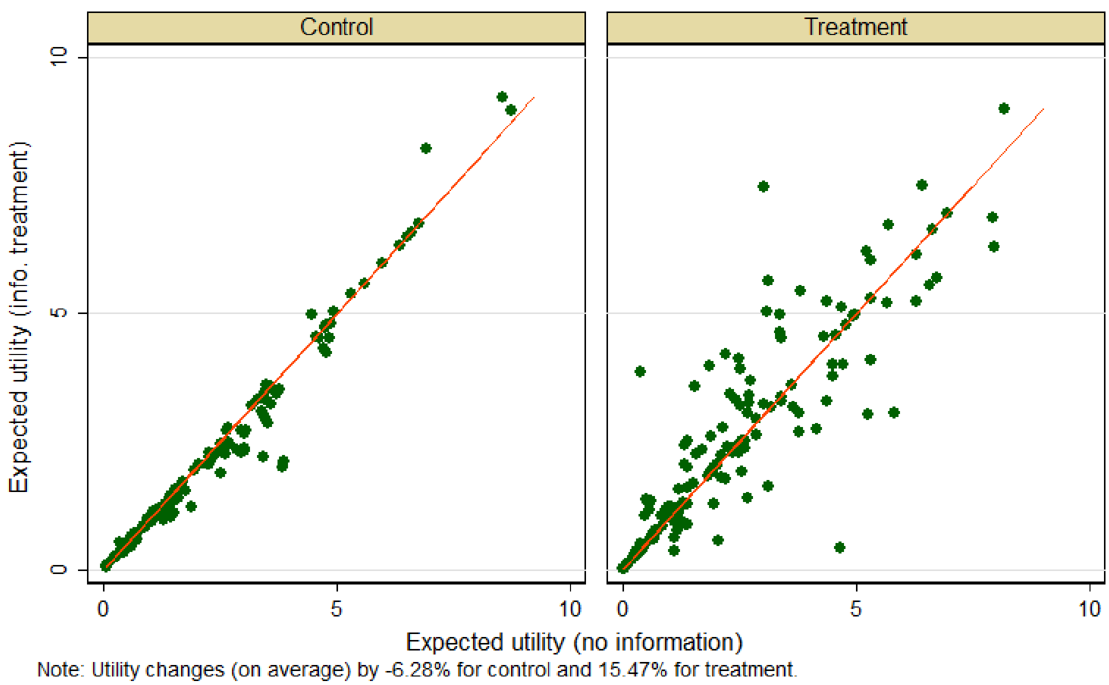

Table 1: Description of the sample of trainees

	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			
Banking	528	0.26	0.44
BPO	522	0.05	0.21
Retail	530	0.15	0.36
Hospitality	530	0.09	0.29
Health Care	532	0.33	0.47
IT	530	0.08	0.27
Other	516	0.06	0.24
Panel B: Salary and social security			
Salary	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

<u>*Notes:*</u> 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	Whole sample: 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

		Pct. of 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 percentage 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	Hypothetical rank	Random Process	Perfect Knowledge
	(1)	(2)	(3)	(4)	(5)
1.	Rank of manager's top choice	7.2^{***}_{***}	5.44_{***}	5.5	11
2.	Average Rank by trainee	6.76^{***}_{***}	4.78^{***}_{***}	6	10
3.	Most preferred by trainee	9.38^{***}_{***}	7.26^{***}_{***}	8.25	11
4.	Correlation b/w preferences	0.1^{**}_{***}	0.17^{***}_{***}	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 and column (3) reports the average job rank as predicted by the hypothetical preferences. Column (4) calculates the rank as if this process was done randomly. Column (5) 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 (4) and (5) respectively. * p<0.1, ** p<0.05 and *** p<0.01 respectively.

Table 8: Impact on interviews and job characteristics

	Atleast one interview	Number of interviews	No. of interviews (Conditional)	Salary	Location	P.F.
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0188	0.0900	0.135	0.0233	-0.0594	-0.0269
	(0.0529)	(0.0906)	(0.109)	(0.102)	(0.0852)	(0.0699)
N	293	293	149	217	217	217
R^2	0.253	0.330	0.388	0.310	0.239	0.193
Mean of control group	0.500	0.693	1.386	1.010	1.052	0.505
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Batch FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Column (3) reports the number of interviews conditional on getting at least one interview. Salary, location and PF are dummy variables where salary takes the value 0,1,2 for low, medium and high category of salary. Location takes the values 0,1,2 for local, same state and out of state job locations. PF takes values 0 and 1 for no and yes respectively. 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. * $p< 0.1$, ** $p<0.05$ and *** $p< 0.01$ level of significance.

Table 9: Impact on quality of jobs

Interview for:	Best four jobs (Dummy)		Job preference
	OLS	Logit	OLS
	(1)	(2)	(3)
Treatment	0.0868	0.425*	0.323*
	(0.0861)	(0.239)	(0.168)
N	141	141	141
R^2	0.207		0.324
Mean of control group	0.559		-0.225
Individual Controls	Yes	Yes	Yes
Batch FE	Yes	Yes	Yes

Notes: Job preferences have been normalised to have mean 0 and standard deviation of 1. Best four jobs is a dummy variable that takes the value 1 if the interview was among the top four ranked jobs. 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. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 10: Rank ordered logit results

	Salary	Location	Provident Fund
	(1)	(2)	(3)
$\bar{\beta}$	0.186	0.975**	0.114
	(0.459)	(0.430)	(0.374)
Female	-0.284**	-0.224**	-0.282***
	(0.114)	(0.108)	(0.0947)
College	0.0942	-0.0689	0.117
	(0.133)	(0.130)	(0.101)
Urban	0.389*	0.0785	0.491**
	(0.206)	(0.118)	(0.199)
Age	0.0175	-0.0536***	0.00758
	(0.0225)	(0.0202)	(0.0178)

Notes: All outcome variables have discrete values ranging from 1-3 for salary and location and 0-1 for PF. Standard errors are clustered at the individual level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 11: Impact on job choice and employment outcomes

	No. of interviews	Offer received	Offer accepted	Same job after 3 months	Same job after 6 months
	(1)	(2)	(3)	(4)	(5)
Treatment	0.107	0.0534	0.0230	0.0191	-0.0083
	(0.0704)	(0.0616)	(0.0348)	(0.0279)	(0.0084)
N	293	293	293	293	266
R^2	0.311	0.138	0.101	0.094	0.062
Mean control group	0.421	0.271	0.086	0.05	0.0078
Ind. Controls	Yes	Yes	Yes	Yes	Yes
Batch FE	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1)-(3) report interview outcomes and columns (4)-(5) report employment outcomes. 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.

Table 12: Impact on placement outcomes using job preferences

				Same job		Any job	
	No. of interviews	Offer received	Offer accepted	3 months	6 months	3 months	6 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Job preference	0.00279	0.000128	-0.00136	-0.00211	-0.00172		
	(0.00855)	(0.00601)	(0.00342)	(0.00248)	(0.00172)		
Treat × Pref.	0.0252**	0.0145*	0.0104**	0.00763**	0.00172		
	(0.0112)	(0.00813)	(0.00502)	(0.00361)	(0.00172)		
Treatment						0.0337	0.0706
						(0.0521)	(0.0530)
N	2417	2417	2417	2417	2201	293	266
R^2	0.114	0.123	0.123	0.119	0.103	0.097	0.076
Mean of control group	0.0511	0.0329	0.0113	0.00607	0.000941	0.250	0.194
Batch FE	No	No	No	No	No	Yes	Yes
Ind. controls	No	No	No	No	No	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	No	No
% increase	49.3%	44.1%	92%	125%	182%	13.48%	36.4%

Notes: All observations are at the trainee-job level in columns (1)-(5) and trainee level in columns (6)-(7). Column (1) is the total number of interviews for a job received by a trainee. Column (2)-(7) are dummy variables that take the value 1 if yes and 0 otherwise. 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.

Table 13: Impact on placement outcomes using index of placement quality

	No. of interviews	Offer received	Offer accepted	Same job after 3 months	Same job after 6 months
	(1)	(2)	(3)	(4)	(5)
Treatment	0.189**	0.107*	0.0865**	0.0611**	0.0142
	(0.0887)	(0.0641)	(0.0410)	(0.0295)	(0.0145)
N	293	293	293	293	266
R^2	0.176	0.179	0.143	0.100	0.062
Ind. Controls	Yes	Yes	Yes	Yes	Yes
Batch FE	Yes	Yes	Yes	Yes	Yes

Notes: 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.

A Appendix figures and tables

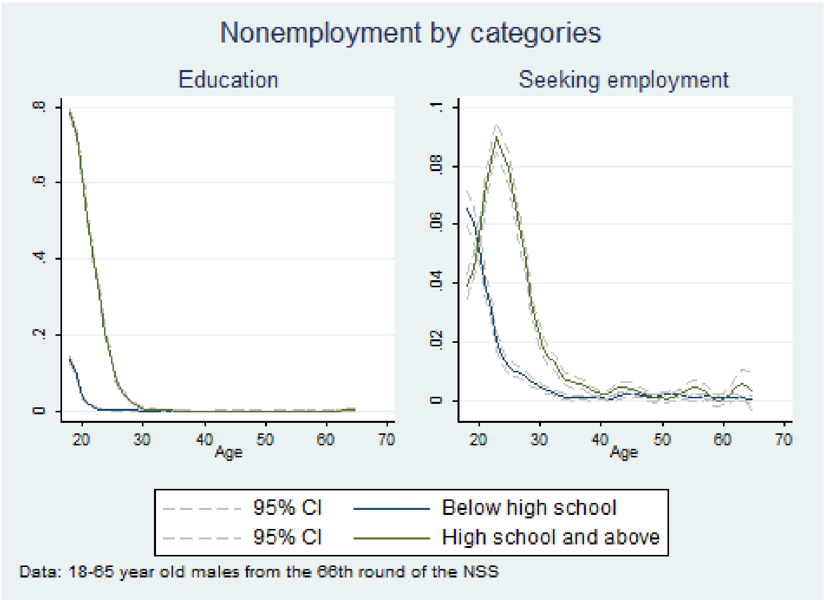

Figure A1: Non-employment rates by categories (2009)

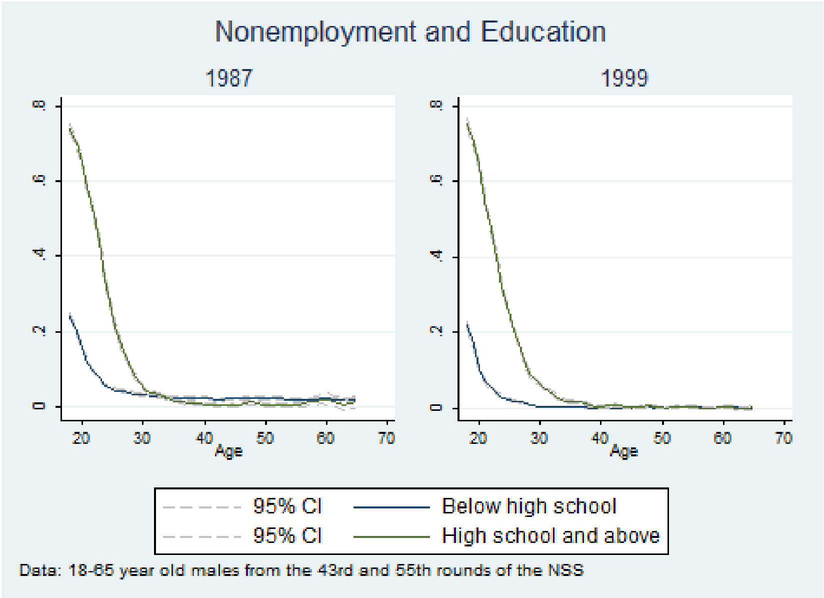

Figure A2: Non-employment rates by education levels (1987 and 1999)

Table A1: 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 A2: 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 A3: Balance check for the intervention

	N	Control	Treatment	p-value
	(1)	(2)	(3)	(4)
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
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).