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

Abhijit Banerjee †   Gaurav Chiplunkar ‡

MIT   Yale University

October 2016

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 a number of 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 close to efficiently matching candidates to interviews subject to their information constraints on preferences. Furthermore, this leads substantial improvement in job choices made by candidates and subsequent employment outcomes for up to six months after initial placement.

*We thank Esther Duflo and Shivakumar Venkatraman for helpful comments. 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

[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.

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 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 into the job placement process in India. It starts by providing detailed evidence for an important source of mismatch in the job placement process of a large provider of job skills training, namely that the placement managers (who are responsible for matching job seekers to interviews) often have no information about the job preferences of the candidates they are responsible for placing and as a result often offer the candidates interviews for jobs the candidates have no interest in.

To document the mismatch, we clearly need to reliably know the preferences of each

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

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 an straightforward thing to do. For reasons explained below, we could not deploy one of the standard incentive compatible mechanisms and in any case, there is evidence that people behave strategically even when truth-telling is the dominant strategy[3]. To elicit the preferences of job seekers over a set of job characteristics, potential employees (who are currently trainees in a job-skills training center) are asked to make choices among some real (and some invented) job options. They are told (and it is in fact true) that they are more likely to get interviews at jobs that they rank higher. 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. Since the two preference distributions we get are essentially identical, we can have 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 the 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 him to pick three best jobs (in order of preference) from that trainee's point of view. This allows us to measure 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

[3] See for example: Engel (2011), Oosterbeek et al. (2004), Palacios-Huerta and Volij (2009), Kagel and Levin (1986), Andreoni (1995), Ferraro and Vossler (2010).

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, 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 alternative assumptions about the what the manager knows–(i) she knows what we know about the preferences of both the treatment *and control* job seekers, (ii) she knows what she tells us about their preferences, (iii) 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 that both assumptions (ii) and (iii) do better than assumption (i) in explaining the data (not surprisingly, managers do not seem to know what we know) and the hybrid case (iii) perhaps fits the data the best. In other words, the managers in this sense, come close to achieving efficiency subject to their information constraints. 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.[4] 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 market matching process, this paper perhaps most directly resembles Pallais (2014) in demonstrating that inexperienced workers suffer from the lack of a track record.

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 under different informational assumptions and what we see in the data. Section 7 reports on the impact of the treatment on job offers, acceptance and retention. We conclude the paper in section 8.

2. Context and Background Data

[4] 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)[5] 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 centre 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 pro-

[5] FICCI Survey on Labour/Skill Shortage for Industry, October 2011.

portion of trainees remain unemployed after the training program. Furthermore, more recent reports from the impact of training programs (NSDC (2013), FICCI (2013)) suggest two major challenges faced by trainers: first, a low take up rate of training programs and second, the tendency of trainees to quit their jobs within a short period (two-three months) of their initial job placement. Both challenges suggest a mismatch between the jobs skilling programs delivery and what their clients want. This could be either because there are not enough of the kinds of jobs the clients want or because the existing pool of jobs are not allocated to the right set of applicants.

For this study, we partner with Skills Academy [6], 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 centre 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

[6]http://theskillsacademy.in

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[7]. As can be seen in column (1), our study sample is young (21 years old on average), have completed their high school education and come from backward caste backgrounds. 48% of the sample is female.

3. Eliciting preferences over jobs

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

3.1 Hypothetical choices

3.1.1 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 (a mandated savings plan). 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

[7] 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.

results for salary and provident fund are reported in panel B of table 2. Trainees report a desired salary of Rs. 15,036 on average[8], with 98% of individuals reporting a preference for a job with provident fund. Panel C reports the location preferences, which is broken down based on the residence of 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.

3.1.2 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

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

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

with real-world job scenarios and discusses what we learn about trainee preferences from their observed choices.

3.2.1 Eliciting unbiased 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

[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.

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 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 centre, 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 amongst the top three jobs (column (4)). We see that there is a substantial heterogeneity in preferences across trainees. For example, 49% of trainees rank job 7 amongst their top three jobs while 46% of jobs rank job 1 amongst their bottom three jobs.

3.2.2 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

[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

the All-India urban CPI for that month.

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

Let us now examine the results: as reported in column (3) of table 5, on average, trainees prefer desk jobs (like receptionists) and desire a modest 5.46% increase in their real salary if they were to work in an active job (like delivery boys) instead. However, this masks a huge heterogeneity in male and female preferences. Males seem to be close to indifferent between desk and active jobs by desiring only a 1.46% increase in real salary in moving to an active job from a desk job. Females on the other hand, strongly dislike active jobs and desire just over a 15% increase in real salary if they have to work in an active job. A similar trend can be seen with respect to working in jobs within the same state of the trainee's residence. On average, trainees are almost indifferent between working in their local area of residence or moving to bigger cities within the state (2.57% increase in real salary). But this masks a huge heterogeneity across gender– males are almost indifferent to moving to bigger cities within the state (0.73% increase in real salary), but females desire a 18.66% increase in the real salary to make them indifferent between working in the district and moving to bigger cities within the state. It is important to take the above interpretations with caution because the regression coefficients are noisy and statistically insignificant, even though the point estimates are big in some cases.

We then examine preferences in terms of relocating outside the state. There is a very strong dislike across trainees to relocate to other cities in India. On average, trainees require a 76.77% increase in their real salary to relocate outside of their home state. Furthermore, like with other characteristics, this masks a stark heterogeneity across gender – males require a 54.48% increase in salary and females desire almost a 136% increase in real salary to be indifferent between staying in their home district and relocating to cities outside their home state. Lastly, with respect to provident fund, both males and females would like to give up around 15% of their real salary to have access to a job with provident fund.

Put together, the above results indicate that trainees (i) prefer more desk/phone jobs as compared to activity intensive jobs; (ii) prefer jobs that provide social security benefits (like provident fund); (iii) are almost indifferent to move to cities within their home state, but have a strong dislike to moving to other cities in the rest of the country. Moreover, we observe large variation in preferences across males and females. It appears to be that males are more likely to take up any job (desk or active) and are more likely to move across locations for the jobs than females, who have a strong preference especially for the job location, which is not surprising, given the Indian context.

3.3 Are the two sets of preferences consistent?

In the above sections, we have described two methods (one based 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. This concurs with and adds to the findings of Delavande et al. (2011) who measure subjective expectations in developing countries, and find that respondents can generally understand and answer probabilistic questions and their expectations are good predictors of future behavior and economic decisions.

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.

4.1 Activity Details

To begin, we first examine if managers are aware of the preferences of trainees. 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, we ask managers to list (in order of preference) three jobs out of the 11 jobs that she would recommend for the trainee. 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 the manager knows her trainee's preferences (described below) and for each of these measures, we report the average rank given by the trainees and compare it to two hypothetical scenarios: one where the manager responded with a random set 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. Measure #1: We consider the 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: For the second measure, we take all the three jobs chosen by the center manager and report the average rank given by the trainee for these jobs. This measure therefore gives us an idea of how good the manager is at knowing the preferences of the trainee on average. As reported in row (2) of table 7, random choice would have generated an average rank of approximately 6 while in the perfect information case it should close to 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: The third measure is 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: Lastly, we take 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 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. He does do slightly better than choosing completely at random, but is nowhere near perfect information. Lastly, 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% of the cases, the manager is able to almost perfectly match the preferences of the trainee (correlation coefficient of 0.9 or more) while in 15.9% cases however, there is almost perfect negative correlation between the choices of the manager and those of the trainee (correlation coefficient of -0.9 or less).

5. The Impact of Informing the Managers

After eliciting the true 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 center managers about the job preferences of the individuals they are in charge of placing and the consequences it had.

5.1 The intervention

The intervention was as follows: trainees in each cohort were randomized into two groups: for the first group (henceforth the Treatment group), we take the job rankings provided by the trainee and send the description of the job characteristics for the top four jobs to the manager. For the second group (henceforth the Control group), no preferences of trainees were shared with the center manager. In figure 5, we list an example with information about two such profiles that were presented to the centre manager. Table A3 checks for the 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 the 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:

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

where T_i is a dummy variable that takes the value 1 for if the trainee was in the treatment group and 0 for the control group. X_i are a set of trainee characteristics like age, gender, education and dummy variables for if the trainee is a student and of lower caste; α_c and α_t are centre and trade fixed effects respectively. 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 are now dummy variables 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 Matching to better quality of interviews?

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 centre managers did a better job at matching the job seekers to interviews that they preferred (as measured by our estimates of their preferences), by reallocating the interviews (rather than by giving them better interviews across the board). 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 did not specify the type of job that they would actually place the trainee in and so we cannot match up this dimension of preferences with the data. We therefore take the 11 jobs and average the rank over the designation dimension. This leaves us with 8 jobs for every trainee that now only vary in terms of salary, location and provident fund.

The bigger challenge was that if we 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 jobs where the set of job characteristics do not map into the jobs provided in the job ranking activity. 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, section 5.4 reports the results for the 141 interviews that we were able to perfectly able to match to trainee preferences and therefore were able to give it a match score. In section 5.5, we

[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%.

then extend our analysis by predicting the preferences for interviews that were not a part of the job ranking list and redo our analysis using the predicted match score for the complete set of 217 interviews. We use two methods for predicting preferences: in subsection 5.5.1, we use a LASSO method to predict rankings based on the reported ranking, while in subsection 5.5.2, we instead use their stated preference weights as reported in the hypothetical exercise (discussed in section 3.1) to generate hypothetical preference rankings.

5.4 The impact on match quality using exact job matches

In this section, we only consider the subset of interviews for which we are able to match the job characteristics to those in the job ranking list and hence we know the preference ranking of the trainee for that interview. We are able to match 141 interviews (65%) of the total 217 interviews. 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.

5.4.1 Treatment Effects

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 provided was in the four most preferred jobs of the trainee and (ii) the (normalized) rank[13] for the interview as reported by the trainee in the ranking exercise. For both the outcome variables, we then estimate the following OLS regression:

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

The results are reported in columns (1) and (3) of table 9. As can be seen in column (1), the provision of information to managers did have a positive effect of matching

[13] 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.

trainees to jobs that were more preferred. A trainee in the treatment group was 15.2 percentage points (or 32.1%) more likely to get an interview for a job that was among her top four preferred jobs. As reported in column (3), trainees in the treatment group interviewed for jobs that were on average 0.35 standard deviations more preferred than those by the control group.

As an additional exercise, since the first outcome variable is a dummy and the second variable is a rank, we estimate a logit and ordered logit specifications respectively with the same set of regressors for both the outcome variables. The results are reported in columns (2) and (4) of table 9. The results are positive, highly significant and consistent with the OLS results reported in columns (1) and (3).

5.4.2 Heterogeneity by manager's knowledge of trainee preferences

As discussed in section 4.1 previously, managers were also asked to choose three jobs (from the same list of 11 jobs) that would be best suited for the trainee and we had constructed four measures to examine how well a manager knew the preferences of the trainee. As reported in the histograms in figure 6, there is considerable variation in the knowledge of the manager across trainees in knowing their preferences. We can therefore examine if the treatment effects are heterogeneous by ex-ante how well the manager knew the preferences of the trainee. Therefore, for each of those four measures, we generate a dummy variable that takes the value 1 if the manager's knowledge (according to that measure) is above the median and 0 otherwise. We then examine the heterogeneity of the treatment effects on both the outcome variables described in the previous section by estimating the following specification:

$$y_{ij} = \beta T_i + \gamma D_i + \delta T_i \text{ x } D_i + \alpha_c + \alpha_t + \varepsilon_{ij} \tag{4}$$

where now, D_i is the dummy variable for a above median (high) knowledge of the trainee's preferences. The results are reported in table 10 [14]. Each column reports the results

[14] Note that the sample size reduces from 141 interviews to 113 interviews. This is because the activity of asking managers to recommend three jobs for trainees as described in section 4.1 was implemented in the last 14 batches only (thus the first 8 batches in our study did not have managers recommend jobs for trainees). Hence the heterogeneity of

when the *high knowledge* dummy is created using that measure. There is no heterogeneity in the impact of the treatment based on the manager's knowledge of trainee preferences.

5.5 Robustness Check: Matching all interviews

As noted earlier, we were only able to perfectly match 65% of the interviews with one of the jobs ranked by the trainees in the job ranking exercise. To be able to match all the job interviews, we predict the preferences for the set of job characteristic combinations that were not in the original list of jobs that trainees ranked. We use two methods to do this: a LASSO method to select the combination of job characteristics that are most predictive of the job preferences and second, use the preference weights given to different job characteristics (as discussed in section 5.5.2) to generate hypothetical preferences.

5.5.1 Using LASSO to predict preferences

One way to do predict preferences for the set of job characteristics that were not in the job ranking list is by letting the machine choose the set of job characteristic combinations that are predictive of the preference for a job. Therefore, we take all the job characteristics along with their first interactions and generate a vector of variables, which we denote by X_j. To isolate which variables are informative in predicting the rank of a job, we run a LASSO estimation by pooling in job rankings for job j across all individuals (denoted by r_{ij}) and regressing them on the entire X_j vector. Therefore, we run the following regression:

$$r_{ij} = \alpha + \beta X_j + \lambda|\beta| + \varepsilon_{ij} \tag{5}$$

The penalty term is chosen such that the LASSO estimation selects the five most predictive variables (call them X_j^s). With these five variables, we then run individual level

the treatment by ex-ante knowledge can only be examined for the interview outcomes of these trainees.

regressions to predict the rank for each job i.e. we run the following OLS specification for every individual i separately:

$$r_j = \alpha + \beta X_j^s + \varepsilon_j \text{ (for every individual } i) \tag{6}$$

where r_j is the rank given by individual i to job j. Using the estimates for $\hat{\alpha}$ and $\hat{\beta}$, we then predict $\hat{r}_j$ for every individual separately. Since we now have the complete ranking across all possible combination of job characteristics, we can match all the 217 interviews based on their job characteristics. We then run the same specification as in equation (3) above, where the dependent variable is now the (normalized) $\hat{r}_{ij}$ instead and report the results in columns (1)-(4) of table 11. As seen in column (1), a trainee in the treatment group is 22.3% more likely (though statistically not significant) to be matched with an interview in his/her top four preferred jobs and as reported in column (3), a trainee is matched to interviews that are on average 0.28 standard deviations more preferred to those in the control group and this is statistically significant. The logit and ordered logit results in columns (2) and (4) concur with these as well and are consistent with those reported earlier in table 9.

5.5.2 Hypothetical Preferences

As a second method, we use the points that trainees had provided (discussed in section 3.1) to weight the job characteristics and generate a hypothetical ranking of jobs. Again, since we now have a complete ranking across all possible combination of job characteristics, we can match all the 217 interviews and run the same specification as in equation (3) above. The results are reported in columns (5)-(8) of table 11. The coefficients on the treatment variable are no longer statistically significant. This result is consistent with the observation that the jobs that managers think trainees want (as reported to us) are not at all consistent with the jobs that trainees would like if we take the hypothetical job rankings as their true preferences (as discussed in section 4.1). Therefore, we should expect no effect of the treatment in matching trainees to better jobs as measured by these hypothetical rankings and indeed the results in table 11 are consistent with this.

6. How does it change overall matching?

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 while saying nothing about the others. This can easily lead the manager to move from an allocation which 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 he chooses the allocation $\{a \rightarrow: 3; b \rightarrow 1; c \rightarrow 2\}$. b gets what manager's best information says should be his top choice. Now suppose the manager is now told very precise information about $a's$ preference and decides that he 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.

A related point is that it is useful to try to evaluate how well the manager is doing in matching people to jobs, given the information he has (though he could gather more information, of course). When we see the manager not giving certain job seekers their most preferred interviews, this could be because everyone wants the same jobs or to put it differently, when we say that in control about 50% get an interview for one of their top four most preferred jobs, how much of this is a reflection of the scarcity of jobs that people want?

Given that in our experiment the treatment and control job seekers were competing for the same pool of interviews, the experiment cannot directly answer these questions.

To make progress, we need to make additional assumptions about the information set of the manager with regard to the preferences of the trainees and ask whether the manager's interview assignment is (or at least comes close to) being in the core given our assumption about her information set.

To do this exercise, we can first assume, counterfactually, that the manager knows the preferences revealed in the job ranking exercise, and check how far the manager's allocation is from what an efficient allocation in the core would predict. Second, we can go to the opposite extreme and base the matching exercise on what the manager thinks are the trainees preferences (as reported by the manager to us),[15] and then check whether she chooses an allocation that is in the core. 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 make the hybrid assumption that 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.

The comparison of the actual allocation with the first benchmark gives us a measure of how well the manager is doing. The comparison of the actual allocation with the second and third benchmark tells us something about whether the manager is using the information we gave her effectively and how far the information moved her choices.

6.1 Stable matching of trainees and interviews

As discussed above, we consider three potential sets of information on trainee preferences that are available to the manager: (i) the job rankings reported by trainees as described in section 3.2; (ii) the jobs reported by centre managers as in section 4.1; (iii) a hybrid ranking generated from the intervention, where we assume that the centre manager knows the revealed preferences for the top four jobs for the treatment group, but only has her guesses (that was reported to us) for the control group. Given these

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

preferences, we run an algorithm that searches for a an efficient allocation of trainees to interviews. An efficient allocation in this context simply implies that once matched to interviews, no two trainees can swap their interviews and be better off.

There are three things related to the algorithm that would be important to clarify here. Firstly, for almost all batches, there are more trainees than interviews– so there will be multiple matching allocations that are in the core. To take this into account, we run the algorithm 25,000 times to simulate stable allocations and thus calculate the probability that a trainee i is matched to an interview for job j. Secondly, a "job" in our setting is purely defined by the salary, location and provident fund characteristics. A variation in any other dimension (work timings for example) is not captured. So empirically, we observe multiple interviews for the same "job". However, since the algorithm matches a trainee to a "job", we can sum the probabilities across the same "job" to calculate the probability that a trainee i is matched to any interview for job j. Thirdly, we observe an individual being matched to multiple interviews. So unless we make further assumptions on how individuals can trade "bundles" of interviews, we cannot compare the theoretical and empirical outcomes. For now, we only consider the batches where less than 15% of the trainees get more than one interview (that we can exactly match to their reported preference). As shown in figure 8, even with this restriction, we are able to examine allocations in 16 out of the 21 batches. In the next section, we do a robustness check where we include all the batches and we show that our qualitative results do not change. Future drafts of the paper will address this concern more carefully.

With these caveats, given the information set of the manager, the matching algorithm generates multiple allocations that are in the core and hence 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 compare this p_{ij} under the three information sets of the manager to what the manager actually does in allocating interviews. In the data, we have a dummy variable (say 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 in figure 9, for low values of p_{ij}, the empirical allocation seems to be quite efficient, irrespective of what we assume to be the manager's information set. On the other hand, there is a stark difference in the allocation efficiency for higher values of p_{ij}. In the second graph of figure 9, we report the density of trainees across p_{ij}. As can be seen, a large density of trainees have a low p_{ij}, which is not surprising given the scarcity of jobs. This makes the interpretation of the first graph unclear.

We therefore consider the area between the curve and the 45 degree line weighted by the mass of trainees at each point of the curve as a measure of inefficiency of the actual allocation. A smaller area would therefore imply lesser inefficiency. In figure 10, for every p_{ij}, we calculate the absolute distance of the curves in the first graph of figure 9 from the 45 degree line and weight it by the density of trainees for that value of p_{ij}. Under the allocations that are predicted using trainee preferences, the actual allocations seem to be very inefficient. However, the actual allocations seem to be more efficient under the allocations predicted using either the manager's preferences or the hybrid preferences. Lastly, there is a stark contrast in efficiency for the cases where a high p_{ij} is predicted. The actual allocations are best explained by efficient allocations predicted using hybrid preferences as opposed to the other two cases.

Lastly, we can find the area under the curves in figure 10 to get a measure of the inefficiency of the allocation. The weighted areas are 115.19, 95.45 and 89.42 for our three cases respectively. This provides two insights: first, using trainee preferences does a bad job at predicting how managers allocate interviews, which indicates the information asymmetry in knowledge of preferences between the trainee and the manager. Second, using the preferences that the manager tells us (second benchmark scenario) does a better job at explaining the interview allocation, but the hybrid scenario (where we assume that the manager uses all the information we give him) does best at matching the actual allocation of interviews to trainees. Put together, the above analysis implies that the treatment did improve the allocation efficiency of matching job seekers to jobs by reducing the asymmetry of information on preferences, especially for those jobs that *should* have a high probability of being allocated to a trainee.

6.2 Stable matches using augmented preferences

One limitation of the above exercise (as discussed more elaborately in section 5.3) is that we have to leave out 76 interviews since they did not correspond to any jobs ranked by the trainees/managers. This may potentially explain why the managers do not behave as the theoretical stable matching algorithm says they should. To give ourselves a better chance at matching the data, we augment the preferences by using a simple dominance argument. To elaborate on this exercise, remember that our jobs vary along the salary, location and provident fund dimensions. We begin by assuming that *conditional* on other jobs characteristics remaining the same, trainees prefer jobs with a higher salary and prefer jobs with provident fund than those without. Next, the location of jobs varies between the local area of residence, city/area within the state and large metropolitan cities (like Delhi and Mumbai) that are outside the state. *Conditional* on the salary and PF, we examine (wherever possible in the job ranking list) how rankings change when location changes and thus impute a preference for location.

Under these assumptions, we are able augment the preferences of trainees (and hence the information set of the managers) and examine if managers match trainees to job interviews that dominate the jobs ranked by the trainee. The results are reported in figure 11. The dotted lines represent the results from figure 9 above and the solid lines represent the results when preferences are augmented. As can be seen, augmenting the preferences does little to change the results.

6.3 Stable matching using all batches

As discussed previously, we had excluded the batches where more than 15% of the trainees got more than one interview. In figures 12 and 13, we replicate the above exercise and instead use all the batches instead of restricting our sample. We see that the qualitative results and interpretations do not change at all. The weighted areas are 175.31, 143.12 and 129.21 for our three cases respectively. Future drafts of the paper will address the issue of multiple interviews more carefully.

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. In this section, we explore whether this further resulted in improving placement outcomes in terms of trainees getting job offers, accepting them and staying employed in the job for up to six months after the completion of the training program.

To begin, we can use the job characteristics of an interview received by the trainee and (potentially) match it to a job in the job list that trainees ranked (as discussed section 3.2). Therefore, for every trainee-job pair we now have: (a) the reported preference by the trainee for that job; (b) various placement outcomes based on the interview(s) that the trainee got for that job[16]. We can then examine the impact of our treatment on the quality of various placement outcomes. It is important to refer back to the discussion in 5.5 due to the fact that there were interviews for jobs that were not in the job list that trainees ranked and hence we do not know their preferences for these jobs. So in section 7.1, we shall examine the impact of our intervention using only the 141 interviews that we can match exactly to one of the jobs in the ranking exercise. In section 7.2, we shall redo the analysis, but now use the predicted preferences from the LASSO method (as discussed in section 5.5.1) to match all the 217 interviews.

7.1 Impact using exact interview matches

[16]For the jobs that the trainee had ranked, but got no interview, we set all outcome variables to 0.

7.1.1 Treatment effects

We begin by taking the 141 interviews that we can exactly match to a trainee-job pair in the job ranking exercise. So for a trainee i and job j, we can estimate the following regression specification:

$$y_{ij} = \alpha_i + \beta P_{ij} + \delta T_i \text{ x } P_{ij} + \varepsilon_{ij} \tag{7}$$

where: y_{ij} are a set of job choices 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, P_{ij} is the job rank as reported by trainee i for job j in the job ranking exercise and is normalised to have mean 0 and standard deviation 1 across all trainees. 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 area estimated using the variation in preferences across jobs within a trainee.

In fact, this exercise is very similar to the one discussed in section 5.4.1, where we had examined if the intervention resulted in an interview being allocated to a trainee that preferred it more. Here we look at the same question, but from the point of view of the trainee by taking all trainee-job pairs and examining the impact of our intervention on not only matching to better interviews, but subsequently affecting other outcome variables as well.

Table 12 reports the results for the above regression specification. We consider three outcomes related to interviews and offers and four outcomes variables related to job retention and employment. First, we examine the impact on the number of interviews. As reported in column (1), relative to the control group, trainees in the treatment group got 0.0254 additional interviews for jobs that were ranked one standard deviation higher. Considering a benchmark in the control group where the average number of interviews was 0.051 (across all trainee-job pairs), this translates into a 49.7% increase in number of interviews. This result concurs with our previous analysis in section 5.4.1 that the intervention matched treatment trainees to interviews for jobs that they prefer more.

Second, we examine whether this had an impact on the number of 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 42.8% 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 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 in the control group.

In columns (4)-(7) of table 12, we examine the impact of our treatment on job retention an employment outcomes. Given that we had two rounds of follow up surveys: the first one after three months and the second one after six months after the completion of the training program, we can examine the impact of our treatment on employment in the short run (three months) and a slightly longer time horizon of six months[17]. In columns (4) and (5), the outcome variables are dummies that take the value 1 if the trainee was employed in the *same* job (that they were matched to) three months and six months later respectively and 0 otherwise. As reported in column (4), trainees in the treatment group were 0.96 percentage points more likely to be employed in jobs that they ranked one standard deviation higher as compared to the control. This translates into just over a 100% increase in the retention of a trainee in the job as compared to the average in the control group. Column (5) looks at job retention after six months and we do not find trainees retaining the same job in the longer run.

This leads us to examine whether trainees were employed in *any* job to take into account job transitions and quits. So in columns (6) and (7), the outcome variables are dummies that takes the value 1 if the trainee was employed in *any* job after three

[17] Note that the number of observations for the 6-month outcome variables are lesser than the 3-month ones because from the trainees we surveyed after three months, we were only able to follow up with 90% of them after six months.

months and six months respectively and 0 otherwise. As reported in column (6), trainees in the treatment group were 1.7 percentage points more likely to be employed three months after they were initially placed in a job they ranked one standard deviation higher as compared to the control group. This translates into a 89% increase in the probability of the trainee being employed as compared to the average in the control group. As reported in column (7), trainees in the treatment group were 1.4 percentage points more likely to be employed in a job six months after they were initially placed in a job that they ranked one standard deviation higher as compared to the control. This translates into a large 253% increase in the probability a trainee is employed in a job six months later.

Lastly, we can redo the analysis for all our outcomes variables using only the most preferred job of the trainee. Therefore, we can examine if providing information on preferences to the manager (through better matching) had an impact on job choice and employment outcomes for the trainee's most preferred job. We report the results in Panel A of table 13. We see that the treatment does not have a differential impact on treatment trainees getting more number of interviews for their best job or the number of offers. However, they are 3.4 percentage points more likely to accept an offer and 3.57 percentage points more likely to stay in the same job three months after placement. Though not more likely to stay in their most preferred job after six months, trainees are 3.65 percentage points more likely to be employed in any job conditional on being matched to their most preferred job. These translate into very large treatment effects given the low mean of the outcome variables for trainees in the control group.

7.1.2 Impact using a placement quality index

Another way of examining the impact of the intervention at the trainee level is to create a "placement quality index" that measures the quality of the placement for each trainee by weighting the outcome by the preference of the trainee for that job. To elaborate, for each trainee i and job j, we can weight the outcome variable y_{ij} with trainee i's (normalised) rank for that job j (i.e. P_{ij}) and sum it across all jobs to create a placement quality index for that outcome for trainee i (denoted by Q_i^y). Therefore:

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

We can then examine the impact of the treatment on this index (by using Q_i^y as the dependent variable). Therefore, for each of the job choice and employment outcomes discussed above, we estimate the following specification:

$$Q_i^y = \beta T_i + \gamma X_i + \alpha_c + \alpha_t + \varepsilon_{ij} \tag{8}$$

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[18]. The results for equation (8) are reported in Panel B of table 13 and concur with the 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 (three months) as well as the long term (six months).

7.2 Treatment effects using all jobs

As noted in the discussion in section 5.5 above, trainees did receive interviews for jobs that were not ranked by them in the job ranking exercise. Since, we have no measure of trainee preferences for these jobs, we have to exclude them from the analysis in the previous section. However, as discussed in section 5.5.1, we can use the reported preferences to predict preferences for all jobs using the LASSO method. Since we now have trainee preferences across *all* job characteristic combinations, we can match *all* interviews to the trainee-job pairs and redo that analysis from the previous section i.e. we can re-estimate equations (7) and (8), but now use the (normalised) *predicted* preferences as P_{ij} instead of the reported ones. The results are reported in 14 and table 15. As seen from both the tables, the impact of the treatment is similar to the previous section on all placement outcomes, though the statistical significance of the coefficients varies

[18] In appendix table A4, we look at the impact of the treatment on simply the outcome variables, ignoring any quality dimension of that outcome. As reported in the table, we see no impact of the treatment on improving outcomes.

due to estimation error in predicting preferences.

8. Conclusion

This paper identifies an important potential source of mismatch in the Indian labor market – that intermediaries (centre managers in our context) who are responsible for matching job seekers to jobs do now 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.

We begin carefully elicit the preferences of trainees over a set of jobs and examine the extent to which there is an asymmetry of information about these preferences between the trainees and their centre managers. Having documented this friction, we provide center managers with this information for a randomly selected group of trainees. We find that providing this information changes the allocation of interviews in the direction of getting better interviews and better jobs for those trainees for whom this information was provided.

However, we cannot infer from this that there was a welfare improvement since there was a reallocation of jobs for whom there was information to those for whom there was no information. To look at this more carefully, we make different assumptions on the information sets of the manager and examine how the actual allocation of interviews compare with the efficient allocations that a matching algorithm would predict under each of these information sets. We find that using trainee preferences does a poor job at predicting actual allocations However, interviews come closest to matching the efficient allocations under which the centre manager uses the information that we provide him, indicating that the centre managers come close to allocating interviews efficiently, subject to their information constraints.

Lastly, we examine the impact of our intervention on job choices and employment outcomes. We find that providing information on preferences does result in trainees being matched to interviews that they prefer more and this has a large impact on job choice and subsequent retention in the same job for up to three months and in any employment for up to six months after the completion of the training program.

References

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.

Andreoni, James, “Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments,” *The Quarterly Journal of Economics*, 1995, *110* (1), 1–21.

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.

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.

Delavande, Adeline, Xavier Giné, and David McKenzie, “Measuring subjective expectations in developing countries: A critical review and new evidence,” *Journal of Development Economics*, 2011, *94* (2), 151–163.

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

Engel, Christoph, "Dictator games: A meta study," *Experimental Economics*, 2011, *14* (4), 583–610.

Ferraro, Paul J and Christian A Vossler, "The source and significance of confusion in public goods experiments," *The BE Journal of Economic Analysis & Policy*, 2010, *10* (1).

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

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

Kagel, John H and Dan Levin, "The winner's curse and public information in common value auctions," *The American economic review*, 1986, pp. 894–920.

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.

Oosterbeek, Hessel, Randolph Sloof, and Gijs Van De Kuilen, "Cultural differences in ultimatum game experiments: Evidence from a meta-analysis," *Experimental Economics*, 2004, *7* (2), 171–188.

Palacios-Huerta, Ignacio and Oscar Volij, "Field centipedes," *The American Economic Review*, 2009, *99* (4), 1619–1635.

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

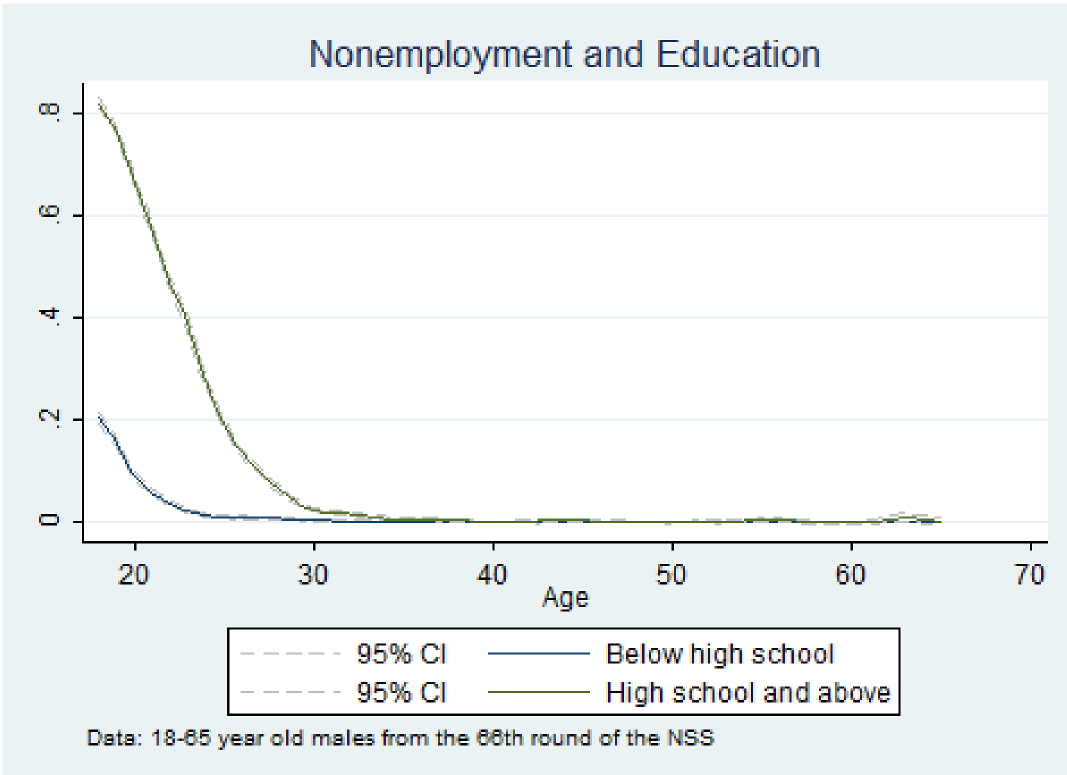


Figure 2: Variation in job characteristics

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:

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

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

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

लखनऊ में एक Team Member/Brew Master का पद मौजूद है। एक Team Member/Brew Master की नौकरी के रूप में आपकी जिम्मेदारियों होंगी – 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

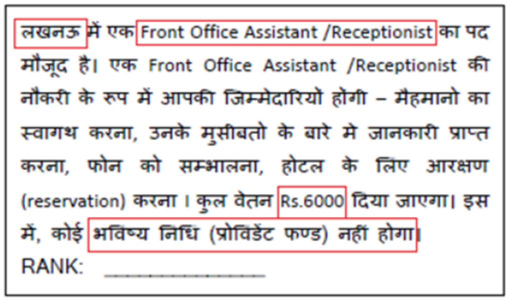
Jaipur में एक Team Member/Brew Master का पद मौजूद है। एक Team Member/Brew Master की नौकरी के रूप में आपकी जिम्मेदारियों होंगी – फ्रंट डेस्क पे मेहमान को संभालना, उनका खाने-पीने का आर्डर लेना, कॉफी बनाना और परोसना। कुल वेतन Rs.6,000 दिया इस में, कोई भविष्य निधि (प्रोविडेंट फण्ड) नहीं होगा।
Rank: ________________

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

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

Malihabad में एक Front Office Assistant /Receptionist का पद मौजूद है। एक Front Office Assistant /Receptionist की नौकरी के रूप में आपकी जिम्मेदारियों होंगी - 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 जिम्मेदारियो होंगी – मैहमानो 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

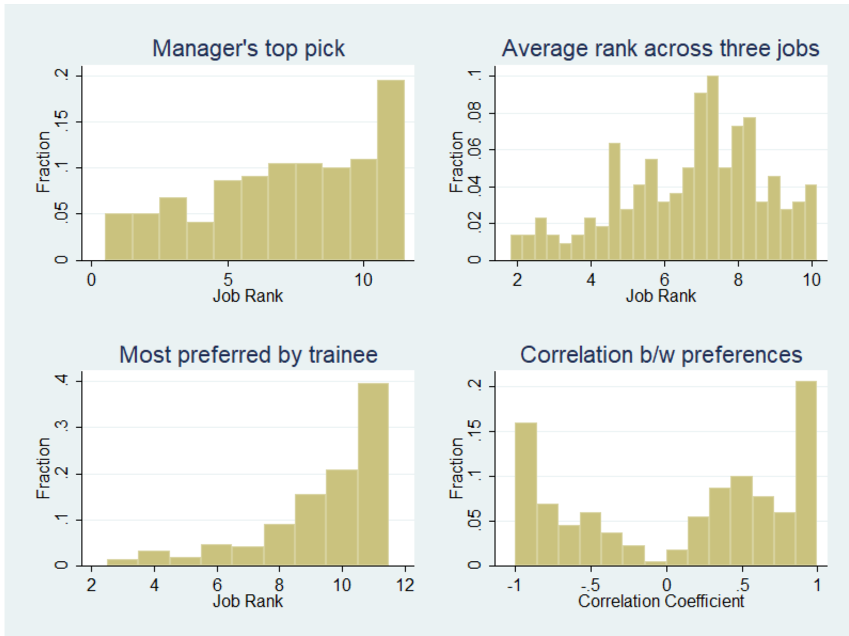


Figure 7: Manager's knowledge of hypothetical trainee preferences

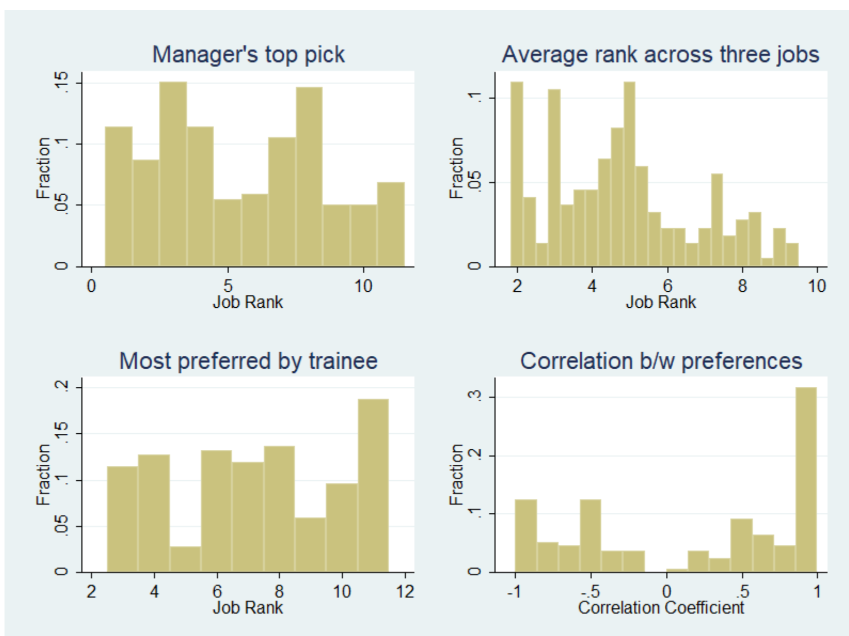


Figure 8: Distribution of trainees and interviews

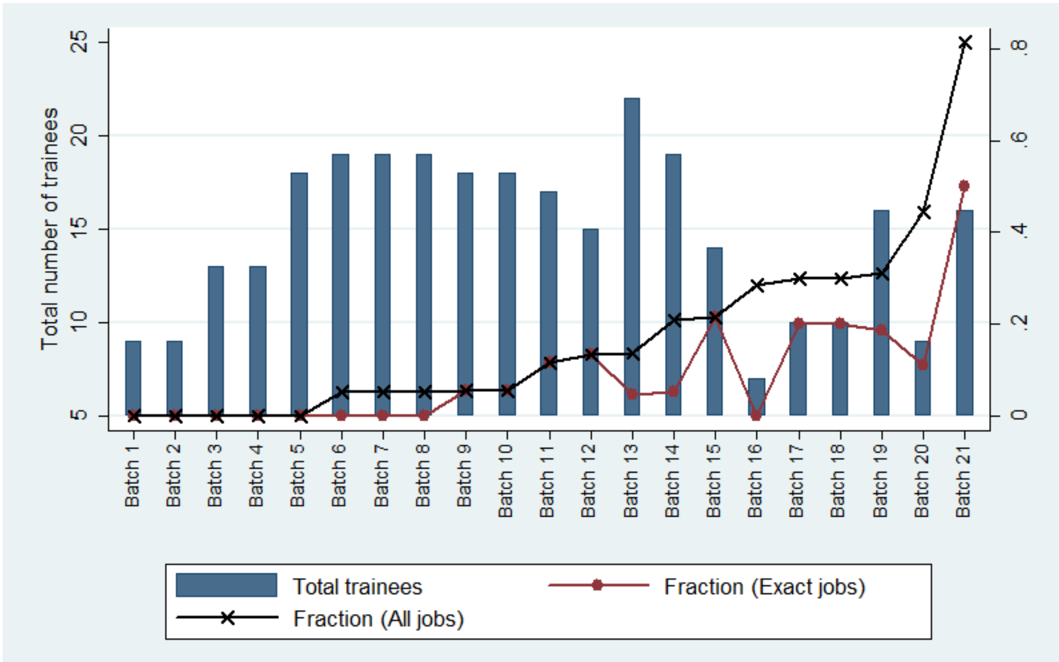


Figure 9: Stable matches and actual outcomes

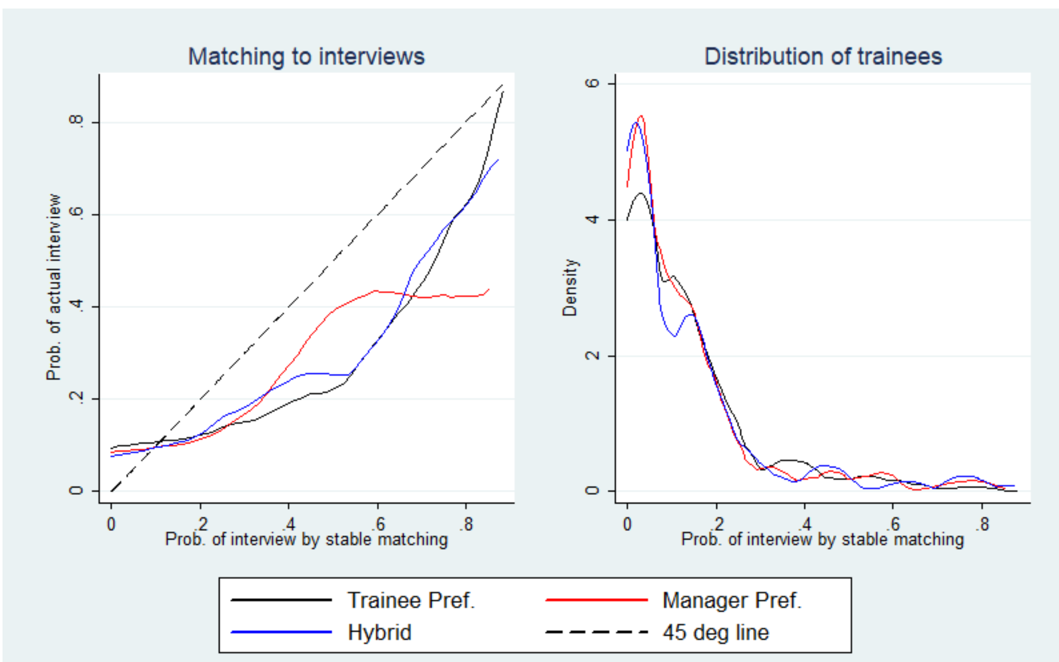


Figure 10: Matching efficiency

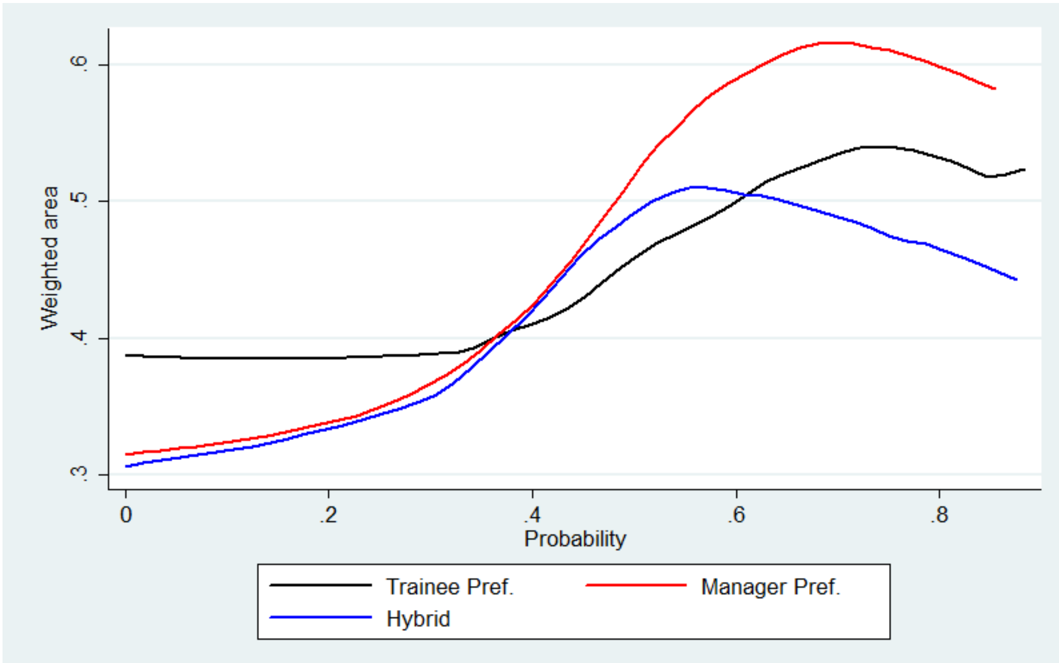


Figure 11: Stable matches with augmented preferences

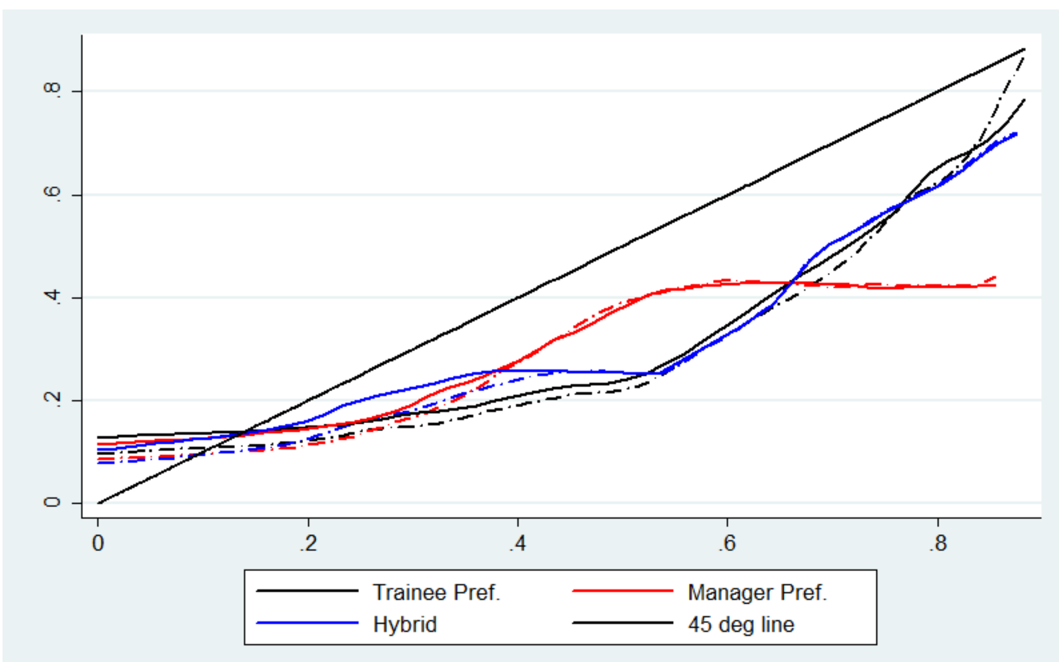


Figure 12: Stable matches (including all batches)

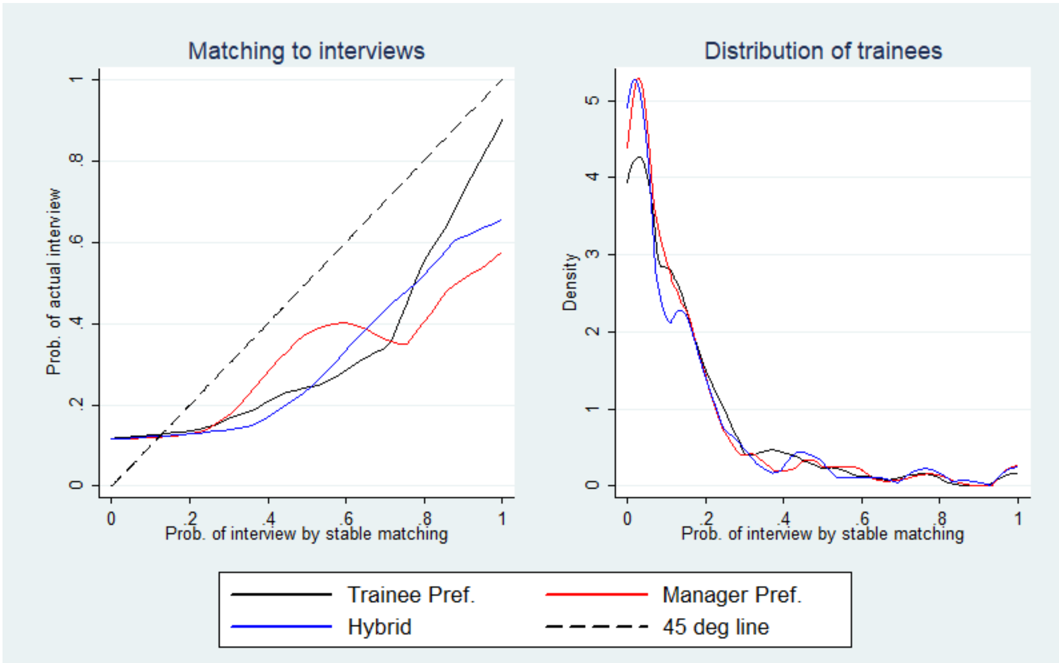


Figure 13: Matching efficiency (with all batches)

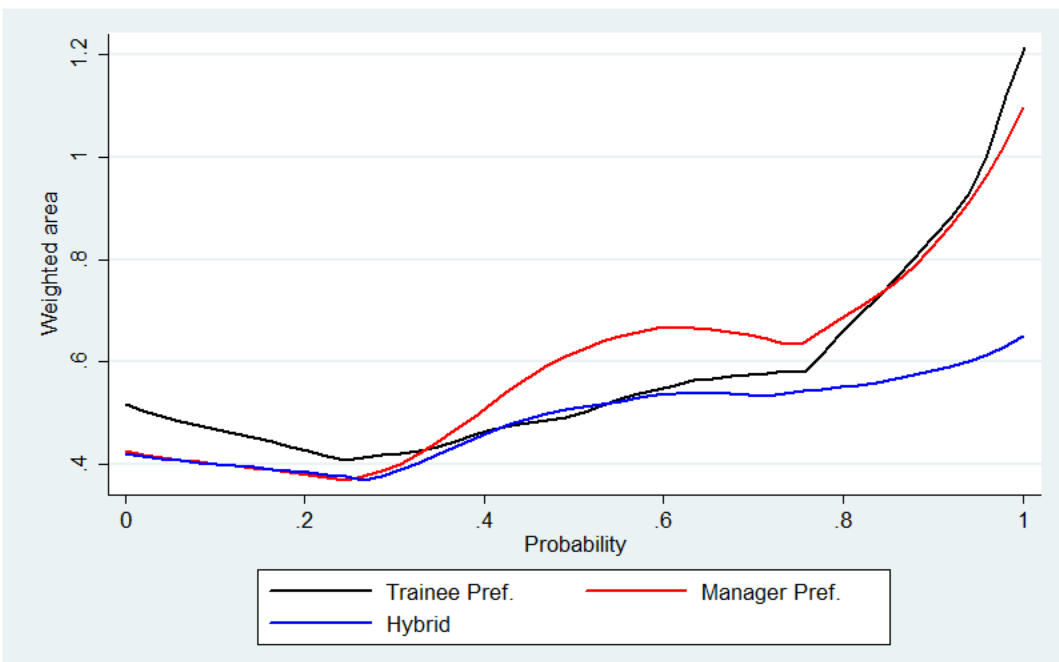


Table 1: Description of the sample of trainees

	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 = 303)	0.18	0.74	0.08
Delhi (N = 67)	0.97	-	0.03

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

Table 3: Distribution of 100 points

Job characteristic	N	Mean	S.D.	Male	Female	p-value
		Whole sample				
	(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	0.40	0.14	4.73	4.38	0.3
Job 2	338	0.39	0.42	0.20	5.45	5.08	0.306
Job 3	338	0.33	0.44	0.23	5.38	5.55	0.616
Job 4	338	0.31	0.49	0.20	5.43	5.61	0.586
Job 5	338	0.12	0.54	0.33	7.05	6.52	0.078
Job 6	338	0.18	0.50	0.31	6.54	6.66	0.716
Job 7	338	0.13	0.38	0.49	7.75	7.71	0.897
Job 8	338	0.32	0.47	0.21	5.31	5.6	0.383
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	0.692
Job 11	289	0.19	0.39	0.42	6.7	7.32	0.106

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***
	(0.0171)	(0.0176)
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. See the paper for details on construction of the hypothetical rank variable. 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

	Reported rank	Hypothetical rank	Random Process	Perfect Knowledge
	(1)	(2)	(3)	(4)
Rank of manager's top choice	7.2^{***}_{***}	5.44_{***}	5.5	11
Average Rank by trainee	6.76^{***}_{***}	4.78^{***}_{***}	6	10
Most preferred by trainee	9.38^{***}_{***}	7.26_{***}	8.25	11
Correlation b/w preferences	0.1^{**}_{***}	0.17^{***}_{***}	0	1

Notes: Each row is a different measure of the manager's knowledge of trainee preferences. The first row reports the average rank for the best job chosen by the manager. The second row reports the average rank across the three jobs chosen by the manager. The third row reports the maximum rank across all the three jobs chosen by the manager. The fourth row reports the correlation between the preferences of the manager and the trainee. Column (1) reports the average job rank as reported by the trainees in the job training exercise as per the measure in the corresponding row and column (2) reports the average job rank as predicted by their hypothetical preferences. Column (3) calculates the rank as if this process was done randomly. Column (4) calculates the rank if the managers were to have perfect knowledge of the preferences of the trainee. The asteriks report the results from a t-test that compare the means in column (1) to column (3) in the top row and column (4) in the bottom row. * $p<0.1$, ** $p<0.05$ 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	0.0203	0.0877	0.129	0.0346	-0.0415	-0.0336
	(0.0412)	(0.0677)	(0.0994)	(0.0955)	(0.119)	(0.0476)
N	293	293	149	217	217	217
R^2	0.205	0.254	0.244	0.301	0.217	0.115
Mean of control group	0.500	0.707	1.414			
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Centre FE	Yes	Yes	Yes	Yes	Yes	Yes
Trade FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Column (1) is a dummy variable that takes the value 1 if the trainee gets at least one interview and 0 otherwise. 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. Standard errors are clustered at the centre level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 9: Impact on quality of jobs

Dependent variable	Best four jobs (Dummy)		Job preference	
	OLS	Logit	OLS	Ordered Logit
	(1)	(2)	(3)	(4)
Treatment	0.152***	0.773***	0.355**	0.776***
	(0.0334)	(0.171)	(0.131)	(0.281)
N	141	141	141	141
R^2	0.217		0.280	
Mean of control group	0.475		-0.203	
Individual Controls	Yes	Yes	Yes	Yes
Centre FE	Yes	Yes	Yes	Yes
Trade FE	Yes	Yes	Yes	Yes

Notes: Job preferences go from 1 to 11 where a larger number implies more preference. 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 provided was among the top four ranked jobs by the trainee 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. Standard errors are clustered at the centre level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 10: Heterogeneous treatment effects

	Best four jobs (Dummy)				Normalised job preference			
	Measure #1	Measure #2	Measure #3	Measure #4	Measure #1	Measure #2	Measure #3	Measure #4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.127	0.269*	0.264	0.124	0.314	0.371	0.467	0.208
	(0.142)	(0.158)	(0.171)	(0.148)	(0.292)	(0.288)	(0.296)	(0.276)
High Knowledge	0.184	0.164	0.184	-0.0203	0.389	-0.00236	0.124	-0.135
	(0.152)	(0.157)	(0.168)	(0.154)	(0.322)	(0.338)	(0.313)	(0.312)
High x Treat	0.138	-0.196	-0.161	0.0643	-0.0154	-0.194	-0.339	0.0697
	(0.199)	(0.211)	(0.209)	(0.198)	(0.392)	(0.419)	(0.395)	(0.369)
N	113	113	113	113	113	113	113	113
R^2	0.381	0.230	0.230	0.222	0.249	0.224	0.227	0.222
Mean of Low x Control	0.429	0.350	0.400	0.474	-0.275	-0.164	-0.0773	-0.0233
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Centre FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable in columns (1)-(4) is a dummy which takes the value 1 if the interview was among the top four ranked jobs by the trainee and 0 otherwise. Columns (5)-(8) is job preferences that go from 1 to 11 where a larger number implies more preference. Job preferences have been normalised to have mean 0 and standard deviation of 1. Four measures of ex-ante knowledge of the manager are considered (see the paper for construction). Each column reports the heterogeneity of the treatment by that measure. High knowledge takes a value 1 if the measure of knowledge was above the median 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. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 11: Impact on job quality (using all jobs)

	LASSO				Hypothetical Preferences			
	Best four jobs (Dummy)		Job preference		Best four jobs (Dummy)		Job preference	
	OLS	Logit	OLS	Ordered Logit	OLS	Logit	OLS	Ordered Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.0550	0.278	0.277**	0.582***	-0.0334	-0.142	0.0399	0.0955
	(0.0721)	(0.452)	(0.119)	(0.161)	(0.0800)	(0.479)	(0.121)	(0.227)
N	217	212	217	217	217	212	217	217
R^2	0.171		0.194		0.173		0.163	
Mean of control group	0.247		-0.147		0.289		-0.0476	
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Centre FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1)-(4) use job preferences that are predicted using LASSO and have been normalised to have mean 0 and standard deviation of 1. Columns (5)-(8) use normalised job preferences generated using weights assigned to a job characteristic in the trainee survey. Best four jobs is a dummy variable that takes the value 1 if the interview provided was among the top four ranked jobs by the trainee 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. Standard errors are clustered at the centre level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 12: Impact on job choice and employment outcomes

	Job choice outcomes			Employment outcomes			
	No. of interviews	No. of offers	Offer accepted	Same job after 3 months	Same job after 6 months	Any job after 3 months	Any job after 6 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Job Preference	0.00281	0.000129	-0.00138	-0.00183	-0.00221	-0.000605	-0.00352
	(0.00861)	(0.00606)	(0.00344)	(0.00303)	(0.00250)	(0.00523)	(0.00283)
Treat. x Pref.	0.0254**	0.0141*	0.0104**	0.00957**	0.00383	0.0170**	0.0143***
	(0.0113)	(0.00834)	(0.00506)	(0.00461)	(0.00364)	(0.00750)	(0.00539)
N	2417	2417	2417	2417	2193	2417	2193
R^2	0.114	0.120	0.123	0.126	0.119	0.152	0.175
Mean of control group	0.0511	0.0329	0.0113	0.00953	0.00376	0.0191	0.00564
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
% increase	49.7%	42.8%	92%	100%	102%	89%	253%

Notes: Columns (1)-(3) report interview outcomes and columns (4)-(7) report employment outcomes. Columns (1)-(2) are the total number of interviews and offers got for a job by a trainee respectively. Column (3)-(7) are dummy variables that take the value 1 if yes and 0 otherwise. Job preferences from from 1 to 11 where a larger number implies more preference. Job preferences have been normalised to have mean 0 and standard deviation 1. Standard errors are clustered at the trainee level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 13: Impact on job and employment outcomes using placement quality and best job

	Job choice outcomes			Employment outcomes			
	No. of interviews	No. of offers	Offer accepted	Same job after 3 months	Same job after 6 months	Any job after 3 months	Any job after 6 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Impact for the most preferred job							
Treatment	0.0536	0.0349	0.0347*	0.0357**	0.00863	0.0404*	0.0365**
	(0.0344)	(0.0259)	(0.0180)	(0.0156)	(0.00825)	(0.0244)	(0.0162)
N	309	309	309	309	280	309	280
R^2	0.118	0.065	0.073	0.083	0.111	0.090	0.071
Mean of control group	0.0612	0.0340	0.00680	0	0	0.0272	0
PANEL B: Impact using the placement quality index							
Treatment	0.188**	0.100	0.0860*	0.0800**	0.0395*	0.137	0.120**
	(0.0797)	(0.0814)	(0.0379)	(0.0343)	(0.0191)	(0.0761)	(0.0387)
N	293	293	293	293	265	293	265
R^2	0.116	0.063	0.092	0.094	0.076	0.124	0.098
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Centre FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Panel A reports the results for only the trainee's most preferred job using the reported preferences. Panel B generates the placement quality index using preferences reported in the job ranking exercise. See the paper for construction of the dependent variables and the placment quality index. 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. Standard errors are clustered at the centre level. * $p< 0.1$, ** $p<0.05$ and *** $p< 0.01$ level of significance.

Table 14: Impact on job choice and employment outcomes (using all job interviews)

	Job choice outcomes			Employment outcomes			
	No. of interviews	No. of offers	Offer accepted	Same job after 3 months	Same job after 6 months	Any job after 3 months	Any job after 6 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Job Preference	0.000729	0.00101	-0.00118	-0.00130	-0.000786	-0.00118	-0.00199
	(0.00491)	(0.00292)	(0.00181)	(0.00166)	(0.000949)	(0.00239)	(0.00121)
Treatment x Preference	0.0130*	0.00495	0.00487*	0.00439*	0.000921	0.00811**	0.00666**
	(0.00676)	(0.00442)	(0.00289)	(0.00263)	(0.00181)	(0.00395)	(0.00287)
N	5274	5274	5274	5274	4770	5274	4770
R^2	0.055	0.058	0.049	0.050	0.052	0.082	0.095
Mean of control group	0.0511	0.0329	0.0113	0.00953	0.00376	0.0191	0.00564
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1)-(3) report interview outcomes and columns (4)-(7) report employment outcomes. Columns (1)-(2) are the total number of interviews and offers got for a job by a trainee respectively. Column (3)-(7) are dummy variables that take the value 1 if yes and 0 otherwise. Job preferences from from 1 to 11 where a larger number implies more preference. Job preferences have been normalised to have mean 0 and standard deviation 1. Standard errors are clustered at the trainee level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Table 15: Impact on job and employment outcomes using placement quality and best job (for all interviews)

	Job choice outcomes			Employment outcomes			
	No. of interviews	No. of offers	Offer accepted	Same job after 3 months	Same job after 6 months	Any job after 3 months	Any job after 6 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Impact for the most preferred job							
Treatment	0.0225	-0.00232	0.000760	0.00302	0	0.00268	0.0133*
	(0.0227)	(0.0166)	(0.00975)	(0.00811)	(.)	(0.0163)	(0.00756)
N	534	534	534	534	482	534	482
R^2	0.082	0.025	0.030	0.029	.	0.054	0.029
Mean of control group	0.0769	0.0577	0.0288	0.0192	0	0.0577	0
PANEL B: Impact using the placement quality index							
Treatment	0.190*	0.0425	0.0516	0.0553	0.0120	0.131*	0.108**
	(0.0912)	(0.0836)	(0.0424)	(0.0393)	(0.0274)	(0.0640)	(0.0408)
N	293	293	293	293	265	293	265
R^2	0.112	0.075	0.071	0.072	0.031	0.095	0.099
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Centre FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Panel A reports the results for only the trainee's most preferred job using preferences predicted by the LASSO method. Panel B generates the placement quality index using preferences generated by LASSO. See the paper for construction of the dependent variables and the placment quality index. 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. Standard errors are clustered at the centre level. * $p< 0.1$, ** $p<0.05$ and *** $p< 0.01$ level of significance.

A. Appendix Figures and Tables

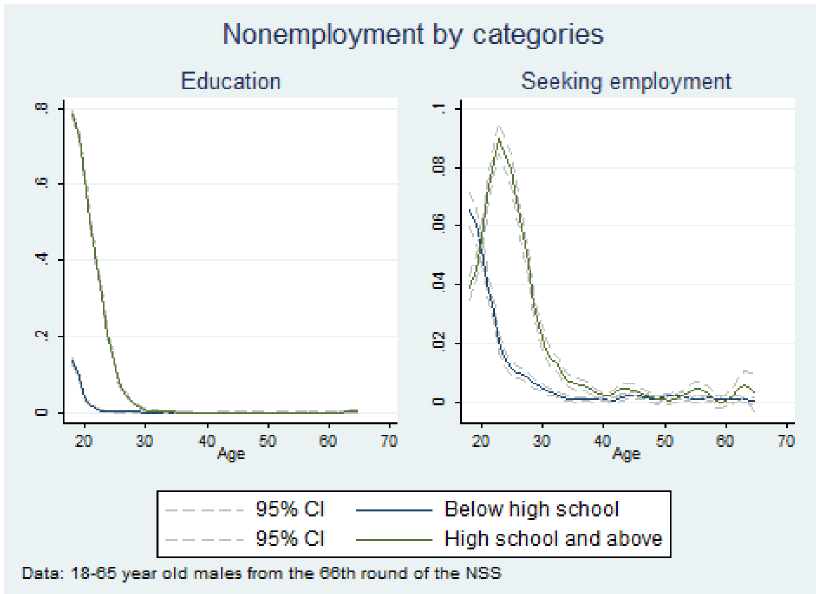


Figure A1: Non-employment rates by categories (2009)

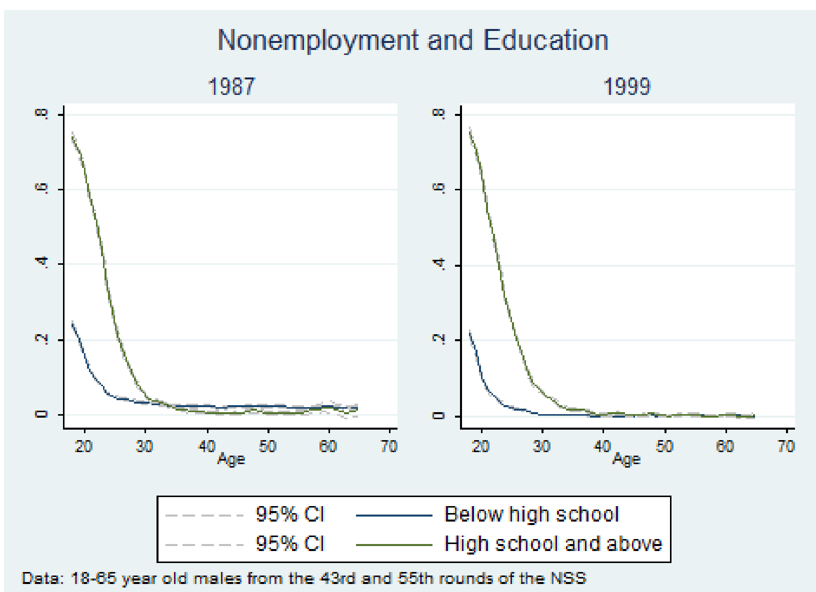


Figure A2: Non-employment rates by education levels (1987 and 1999)

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 contrl 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).

Table A4: Impact on job choice and employment

	No. of interviews	No. of offers	Accept offer?		Stay employed?	
	OLS	OLS	OLS	Logit	OLS	Logit
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0877	0.00336	-0.0154	-0.203	0.0269	0.161
	(0.0677)	(0.0259)	(0.0342)	(0.409)	(0.0371)	(0.214)
No. of interviews		0.593***				
		(0.0668)				
No. of offers			0.301***	2.497***	0.168***	0.859***
			(0.0311)	(0.348)	(0.0264)	(0.129)
N	293	293	293	276	293	293
R^2	0.254	0.578	0.400	0.417	0.131	0.11
Mean control group	0.707	0.443	0.179		0.25	
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Centre FE	Yes	Yes	Yes	Yes	Yes	Yes
Trade FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Accept offer is a dummy that takes the value 1 if the job offer is accepted and 0 otherwise. Stay employed is a dummy that takes the value 1 if the trainee is employed and 0 otherwise. Trainee controls used are the 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. Standard errors are clustered at the centre level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.